

This page Is Inserted by IFW Operations  
And is not part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of  
The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## **IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
Please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**



	10	85.6	5.2	246	24	ABL763561	Corn tassell-derive
	11	62.4	3.8	309	24	ABL71938	corn tassell-derive
	12	42	2.6	8067	24	ABN60224	Human Chemically m
	13	41.8	2.6	10057	22	ABAI8426	Human nervous syst
	14	41.4	2.5	12669	24	ABL70490	Chemically treated
	15	41.4	2.5	12669	24	AAS61449	Human gene regulat
	16	41.4	2.5	12669	24	ABK31523	Signal transductio
	17	40.8	2.5	1887	20	AAK99539	Nucleic acid seque
C	18	40.8	2.5	3221	24	ABO659222	Listeria monocytog
C	19	40.8	2.5	6921	24	ABQ71071	Listeria monocytog
C	20	39.4	2.4	6557	24	ABL33000	Human immune syste
C	21	39.2	2.4	8588	22	AAS45469	Chemically pretrea
C	22	39.2	2.4	8588	24	ABK28325	DNA transcription
C	23	39	2.4	1231	24	AAD22126	Lycopodium escul
C	24	38.8	2.4	2718	20	AAV82021	Moraxella catarrha
C	25	38.8	2.4	14041	22	AAH48024	Internal control B
C	26	38.8	2.4	49617	22	AAF28541	Genomic fragment #
C	27	38.6	2.4	924	24	ABN68768	Streptococcus poly
C	28	38.6	2.4	1176	22	AAHS2921	S. epidermidis ope
C	29	38.6	2.4	1398	24	ABN70725	Streptococcus ope
C	30	38.6	2.4	3310	22	AAH54179	S. epidermidis gen
C	31	38.6	2.4	3332	22	AAH54155	S. epidermidis gen
C	32	38.6	2.4	5642	24	ABL32869	Human immune syste
C	33	38.6	2.4	7695	24	ABL92259	Chemically treated
C	34	38.6	2.4	7695	24	ABL33537	Human immune syste
C	35	38.6	2.4	12601	24	ABL42027	Human immune syste
C	36	38.6	2.4	2155561	24	ABN71527	Streptococcus poly
C	37	38.4	2.3	1269	24	ABN91748	Staphylococcus epi
C	38	38.2	2.3	903	24	ABO68092	Listeria monocytog
C	39	38.2	2.3	903	24	ABO69810	Listeria monocytog
C	40	38.2	2.3	5347	24	ABL33669	Human immune syste
C	41	38.2	2.3	2944528	24	ABA03041	Listeria monocytog
C	42	38	2.3	1005	22	AAH52444	S. epidermidis ope
C	43	38	2.3	1017	24	ABN92855	Staphylococcus epi
C	44	38	2.3	1441	18	AAV75311	Staphylococcus aur
C	45	38	2.3	3058	22	AAH54221	S. epidermidis gen

ALIGNMENTS

```

RESULT_1
AAA30922 ID AAA30922 standard; DNA: 1637 BP.
XX AC AAA30922:
XX DT 22-SEP-2000 (first entry)
XX DE C. pneumoniae ATP/ADP translocase coding sequence.
XX ATG/ADP translocase; Chlamydia Infection; diagnosis; therapy; ds.
OS Chlamydia pneumoniae.
XX CH Key Location/Qualifiers
FH 51..1598 /*tag= a
FT CDS /product= ATP/Adp_translocase
FM
FN WO200039157-A1.
PN 06-JUL-2000.
PD
PE 22-DEC-1999; 99WO-CA01224.
XX PR 28-DEC-1998; 98US-0114060.
PR 12-MAR-1999; 99US-0123967.
PR 30-JUN-1999; 99US-0141271.
XX (CONN-) CONNUGHT LAB LTD.
XX Murdin AD, Oomen RP, Wang J, Dunn P;
XI

```

XX WPI: 2000-452368/39.  
 DR P-SDB: AAY90265.  
 XX Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,  
 PT prevention and treatment of Chlamydia infection in mammals -  
 PS Claim 2: Fig 1: 81bp; English.  
 XX This sequence encodes the Chlamydia pneumoniae ATP/ADP translocase  
 CC the DNA or protein, are useful for diagnosing, preventing or treating  
 CC Chlamydia infection. The sequences can also be used in a method for  
 CC the detection of Chlamydia infection. Primers or probes derived from the  
 CC DNA sequence are useful in diagnostic tests for detecting Chlamydia  
 CC infection.  
 XX  
 XX Sequence 1637 BP; 431 A; 327 C; 325 G; 554 T; 0 other:  
 SO  
 Query Match 100.0%; Score 1637; DB 21; Length 1637;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1637; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 721 TACGCTTTTGATGGCTATGACATATTGATCTGACCTGTTCTTATGACCAGTTACTGCT 780  
 Oy 781 GGATCAATAGACGATATTGACGATCCGCTCTATTAATCCAGAAATGCAAAAG 840  
 Db 781 GGATCAATAGACGATATTGACGATCCGCTCTATTAATCCAGAAATGCAAAAG 840  
 Oy 841 GGAAGAAAGGCTTAACCTTAATGAATATGAAGATAGCTTCTATCTTATAGAT 900  
 Db 841 GGAAGAAAGGCTTAACCTTAATGAATATGAAGATAGCTTCTATCTTATAGAT 900  
 Oy 901 CTCCTTATATCTTTTATTAATCTCTTGGTATTTGCTATGATTTGCTTAA 960  
 Db 901 CTCCTTATATCTTTTATTAATCTCTTGGTATTTGCTATGATTTGCTTAA 960  
 Oy 961 TCGAAGTGACTTGGAAAAAGTACGTAACACTGCAATATCCATATGATATGATATG 1020  
 Db 961 TCGAAGTGACTTGGAAAAAGTACGTAACACTGCAATATCCATATGATATGATATG 1020  
 Oy 1021 AGTTCATGGGGAACTTCTCTTGGACCTGGCTGATATCCGTAATGCTATTTG 1080  
 Db 1021 AGTTCATGGGGAACTTCTCTTGGACCTGGCTGATATCCGTAATGCTATTTG 1080  
 Oy 1081 TTGCTGGTACGCTATTCGTAATTTGGATGTTAACTGAGCCCTAGTCACTCTGCA 1140  
 Db 1081 TTGCTGGTACGCTATTCGTAATTTGGATGTTAACTGAGCCCTAGTCACTCTGCA 1140  
 Oy 1141 TGGTTCTGCTACAGATATGCTTTCTTGGCTTTGTTATCTTATAGAACCAAGCTTCTG 1200  
 Db 1141 TGGTTCTGCTACAGATATGCTTTCTTGGCTTTGTTATCTTATAGAACCAAGCTTCTG 1200  
 Oy 1201 GCGTGTGCTATGTTGGTGAACAACCTCTCATGTAAGCTGTTGGAGCTATAC 1260  
 Db 1201 GCGTGTGCTATGTTGGTGAACAACCTCTCATGTAAGCTGTTGGAGCTATAC 1260  
 Oy 1261 AGAATATCTTTCGAATCCACAATAATACGCTCTTGTGACTCACTAAGAAATGGCCT 1320  
 Db 1261 AGAATATCTTTCGAATCCACAATAATACGCTCTTGTGACTCACTAAGAAATGGCCT 1320  
 Oy 1321 ATATCCCTCTTGACACAGACCAAAATGCAAAAGGTAAGGCTGATATGATGTTGCG 1380  
 Db 1321 ATATCCCTCTTGACACAGACCAAAATGCAAAAGGTAAGGCTGATATGATGTTGCG 1380  
 Oy 1381 CCGGCTTGGAAATTCAGAGAGAGCTTTAATCCACAAGGTTTGCTGTTATCTGTTGAA 1440  
 Db 1381 CCGGCTTGGAAATTCAGAGAGAGCTTTAATCCACAAGGTTTGCTGTTATCTGTTGAA 1440  
 Oy 1441 GTATTGAGCTATGACCCCTTATCTTGCAGTATCTTCTTTCATGATGCTATTTGCT 1500  
 Db 1441 GTATTGAGCTATGACCCCTTATCTTGCAGTATCTTCTTTCATGATGCTATTTGCT 1500  
 Oy 1501 TGGTTCTGCAACTAATGTTAAACAACTATTCTTAGCGAGCTGCTTAAAGAACAG 1560  
 Db 1501 TGGTTCTGCAACTAATGTTAAACAACTATTCTTAGCGAGCTGCTTAAAGAACAG 1560  
 Oy 1561 AAGTGCTCAAGAAATTCAGCTTCGCTTCTTCAATAGAGTGTCTCTTACTCTTGT 1620  
 Db 1561 AAGTGCTCAAGAAATTCAGCTTCGCTTCTTCAATAGAGTGTCTCTTACTCTTGT 1620  
 Oy 1621 GATCCCTACTGCTTT 1637  
 Db 1621 GATCCCTACTGCTTT 1637

RESULT 2  
 AAX91990  
 ID AAX91990 standard; DNA: 1230025 BP.  
 XX AAX91990:  
 AC  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 XX

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.  
 OS Chlamydia pneumoniae.

XX MO9927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-1B01890.

XX 04-NOV-1998; 98US-0107078.

XX 21-NOV-1997; 97FR-0014673.

XX (GSEST ) GENSET.

XX Griffals R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Claim 1; Page 291-611; 1912pp; English.

XX The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.

XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 99.8%; Score 1633.8; DB 20; Length 1230025;

Best Local Similarity 99.9%; Pred. NO. 0;  
 Matches 1635; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAATATAAACTATCAGATAGAAAATTAATAATTTTCAGAGGCTAATAATGACAAAA 60  
 DB 404320 GAAATATAAACTATCAGATAGAAAATTAATAATTTTCAGAGGCTAATAATGACAAAA 404379  
 OY 61 CCGAAGAAAAACCTTTGGAAAAATTCGCTCTTCTGTGGCCGATACATCTCAGAGC 120  
 DB 404380 CCGAAGAAAAACCTTTGGAAAAATTCGCTCTTCTGTGGCCGATACATCTCAGAGC 404439  
 OY 121 TAAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGTATTCATTTAACTATACGGTGT 180  
 DB 404440 TAAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGTATTCATTTAACTATACGGTGT 404499  
 OY 181 TACGGATACAAAAGACACTCTTATTTGGGAGCTCTGTTCTGTGTGTCAGAGGCAATAC 240  
 DB 404500 TACGGATACAAAAGACACTCTTATTTGGGAGCTCTGTTCTGTGTGTCAGAGGCAATAC 404559  
 OY 241 CTTTCATCAAGTTTGGGCTGTGTGTCGCCCTGTGCTATATCTTTATGCTTATTTATGCAA 300  
 DB 404560 CTTTCATCAAGTTTGGGCTGTGTGTCGCCCTGTGCTATATCTTTATGCTTATTTATGCAA 404619  
 OY 301 AGCTAAGTATATTTTAAAGTAAAGAGGCTTATTTATGAGTGGGAAGGCCCTTTTAA 360  
 DB 404620 AGCTAAGTATATTTTAAAGTAAAGAGGCTTATTTATGAGTGGGAAGGCCCTTTTAA 404679  
 OY 361 TTTTCTTGGCCCTGTTCGCCAGTGAATTTATCCGCTACGGGANTTTTACATCCCTACAG 420  
 DB 404680 TTTTCTTGGCCCTGTTCGCCAGTGAATTTATCCGCTACGGGANTTTTACATCCCTACAG 404739  
 OY 421 AATTGGCGACGCTTAAAGGCGCATCCACCCAGAGATGCTAGAGACTCGTGGCANTCT 480  
 DB 404740 AATTGGCGACGCTTAAAGGCGCATCCACCCAGAGATGCTAGAGACTCGTGGCANTCT 404799

OY 481 TAAGAACTGACATTTGCTGCAATTTTAATGACTTGTGAACATATGGGAAAGCGTCATGC 540  
 DB 404800 TAAGAACTGACATTTGCTGCAATTTTAATGACTTGTGAACATATGGGAAAGCGTCATGC 404859  
 OY 541 TATCTCTAATGTTCTGGGGATTTGCTAATGAATAATCAAAAAATCCAGCAAGCAAGCGCT 600  
 DB 404860 TATCTCTAATGTTCTGGGGATTTGCTAATGAATAATCAAAAAATCCAGCAAGCAAGCGCT 404919  
 OY 601 TCTACGCTCTTTTGGGATTCGAGCTAATATTTCTTTACTAGCTTCTGCTGCAATTC 660  
 DB 404920 TCTACGCTCTTTTGGGATTCGAGCTAATATTTCTTTACTAGCTTCTGCTGCAATTC 404979  
 OY 661 TTTGGGCTTCAAAAGTTGAGAGCTTCGTTTGAAGGATGATGATCTTTGGGAAATTCCT 720  
 DB 404980 TTTGGGCTTCAAAAGTTGAGAGCTTCGTTTGAAGGATGATGATCTTTGGGAAATTCCT 405039  
 OY 721 TACGCTCTTTGATGGCTATGACTATTTATCTGACATTTGCTTATGGCCAGTTACTGT 780  
 DB 405040 TACGCTCTTTGATGGCTATGACTATTTATCTGACATTTGCTTATGGCCAGTTACTGT 405099  
 OY 781 GGATCAATTAAGACGCTATTTGACCGATTCCTCGCTTATATTCAGAGCAAGAAATGCAANG 840  
 DB 405100 GGATCAATTAAGACGCTATTTGACCGATTCCTCGCTTATATTCAGAGCAAGAAATGCAANG 405159  
 OY 841 GAAAAAAGGTGCTAACCTTAATAATGAATGAAGATAGCTTCCTATCTATCTGATAGAT 900  
 DB 405160 GAAAAAAGGTGCTAACCTTAATAATGAATGAAGATAGCTTCCTATCTATCTGATAGAT 405219  
 OY 901 CTCTCTTATATCTTTTATTAATCTCTTGTGTTATGCTTATGCTTATTTGCAATTAATCTTA 960  
 DB 405220 CTCTCTTATATCTTTTATTAATCTCTTGTGTTATGCTTATGCTTATTTGCAATTAATCTTA 405279  
 OY 961 TCGAAGTACTTGGAAAAAGTACGCTGAAGACTGCAATATCTAATATGAAATGACTATAGTG 1020  
 DB 405280 TCGAAGTACTTGGAAAAAGTACGCTGAAGACTGCAATATCTAATATGAAATGACTATAGTG 405339  
 OY 1021 AGTTATGAGGGAATTCCTCTCTGACAGGCTGATGCTGATCTTATATCATGATATTTG 1080  
 DB 405340 AGTTATGAGGGAATTCCTCTCTGACAGGCTGATGCTGATCTTATATCATGATATTTG 405399  
 OY 1081 TTGCTGCTAAGCTCAATTCGTAATTTGATGGTAACTGAGAGCCCTAGCTACTCTCTGCA 1140  
 DB 405400 TTGCTGCTAAGCTCAATTCGTAATTTGATGGTAACTGAGAGCCCTAGCTACTCTCTGCA 405459  
 OY 1141 TGGTCTCTCTAAGCTATTCGTTTCTTCTGCTCTTGTATCTTTAAGAAACCAAGCTTCTG 1200  
 DB 405460 TGGTCTCTCTAAGCTATTCGTTTCTTCTGCTCTTGTATCTTTAAGAAACCAAGCTTCTG 405519  
 OY 1201 GCGTGGTGGCTATGTTGGTACAACTCGTCATGCTGAGCTGGTGTGGAGGCTATAC 1260  
 DB 405520 GCGTGGTGGCTATGTTGGTACAACTCGTCATGCTGAGCTGGTGTGGAGGCTATAC 405579  
 OY 1261 AGAATATTTCTTTCGAATTCACAAAATACGCTCTCTTGGACTCAACTAAGAAATAGGCT 1320  
 DB 405580 AGAATATTTCTTTCGAATTCACAAAATACGCTCTCTTGGACTCAACTAAGAAATAGGCT 405639  
 OY 1321 ATATCCCTCTTTCGACCAAGCAAAAAAGTCAAGAGCTGCTATTTGATGATGTTGCCG 1380  
 DB 405640 ATATCCCTCTTTCGACCAAGCAAAAAAGTCAAGAGCTGCTATTTGATGATGTTGCCG 405699  
 OY 1381 CCGGCTTGGGAAATCAGAGAGGCTTATATCCCAAGAGTTGCTGTATTCGTGTGAA 1440  
 DB 405700 CCGGCTTGGGAAATCAGAGAGGCTTATATCCCAAGAGTTGCTGTATTCGTGTGAA 405759  
 OY 1441 GTATTGAGCTATGAGACCCCTTATCTTTCAGAGTATCTCTTTCATCAATTCCTATTTGGT 1500  
 DB 405760 GTATTGAGCTATGAGACCCCTTATCTTTCAGAGTATCTCTTTCATCAATTCCTATTTGGT 405819  
 OY 1501 TGGTTTCTGCAACTTAAGTTAAACAACATATTTCTTACGCACTGTGCTTTTAAAGAACAG 1560  
 DB 405820 TGGTTTCTGCAACTTAAGTTAAACAACATATTTCTTACGCACTGTGCTTTTAAAGAACAG 405879

QY 1561 AAGGCGCTAAGAAGATTACCTCCTCTTTCATAGAGTTGCTTCTTACTCTTGT 1620  
|||||  
Db 405880 AAGGCGCTAAGAAGATTACCTCCTCTTTCATAGAGTTGCTTCTTACTCTTGT 405939  
QY 1621 GATCCCTACCTGCTTTT 1637  
|||||  
Db 405940 GATCCCTACCTGCTTTT 405956

RESULT 3  
AAC81914/c  
ID AAC81914 standard; DNA: 273254 BP.  
XX AAC81914:  
AC AAC81914:  
XX 27-FEB-2001 (first entry)  
DT  
XX Chlamydia pneumoniae genome DNA.  
DE  
XX Chlamydia pneumoniae genome DNA.  
KE Genome: diagnosis; vaccine; ds.  
XX  
XX Chlamydia pneumoniae.  
XX  
PN MO200027994-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 12-NOV-1999; 99MO-US26923.  
XX  
PR 12-NOV-1998; 98US-0108279.  
PR 08-APR-1999; 99US-0128606.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Stephens R, Mitchell W, Kalman S, Davis R;  
XX  
XX WPI: 2000-376516/32.  
DR  
XX  
XX  
PT Isolated nucleic acid for use in diagnostic and analytical methods  
PT encodes genomic sequence of Chlamydia pneumoniae -  
XX  
XX  
PS Claim 2; Page 128-320; 320pp; English.  
XX  
XX This invention describes a novel nucleic acid (N1) encoding a Chlamydia  
CC pneumoniae protein (P1), given in the specification. The isolated nucleic  
CC acid is useful for diagnostic and analytical methods, such as,  
CC hybridization-based assays or amplification-based assays. The protein may  
CC be used for diagnostic purposes, for their enzymatic or structural  
CC activity, or as a vaccine. The invention also describes (1) a probe  
CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid  
CC (N2) that hybridizes under stringent conditions to N1; (3) an expression  
CC cassette comprising N1 under the transcriptional regulation of a  
CC transcriptional initiation region functional in an expression host, and a  
CC transcriptional termination region; (4) a cell comprising an expression  
CC cassette of (3) as part of an extrachromosomal element or integrated into  
CC the genome of a host cell as a result of induction of the expression  
CC cassette into the host cell, and the cellular progeny of the host cell;  
CC (5) a method for producing a P1 comprising growing a cell of (4) where  
CC the protein is expressed and isolating the protein free of other  
CC proteins; (6) a purified polypeptide composition comprising at least 50  
CC weight % of P1; and (7) a monoclonal antibody binding specifically to the  
CC peptide of (6).  
CC  
XX  
SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

Query Match 99.7%; Score 1632.2; DB 21; Length 273254;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAATAAAAAATCATGCAATAGAAAATTAAGTATTTACAGAGGTAATATGACAAAA 60  
|||||  
Db 212470 GAAATAAAAAATCATGCAATAGAAAATTAAGTATTTACAGAGGTAATATGACAAAA 212411

QY 61 CCGAAGAAAAACCTTTGGAAAAATTTGGCTCTTCTTGTGGCCGATACATACGAGC 120  
|||||  
Db 212410 CCGAAGAAAAACCTTTGGAAAAATTTGGCTCTTCTTGTGGCCGATACATACGAGC 212351  
QY 121 TAAAGAAAGTCTGCCAATGTTCCTAATGTCTCTGTATTAACATTAACGGTGT 180  
|||||  
Db 212350 TAAAGAAAGTCTGCCAATGTTCCTAATGTCTCTGTATTAACATTAACGGTGT 212291  
QY 181 TACGCGATACAAAAGACACTCTTATGTGAGAGCTCCGCTTGTGTCAGAGCAATAC 240  
|||||  
Db 212290 TACGCGATACAAAAGACACTCTTATGTGAGAGCTCCGCTTGTGTCAGAGCAATAC 212231  
QY 241 CTTTCATCAAGTTTGGCTTGTGCCCTGTCCCTGTGCTAATTAATCTTATTTATGCAA 300  
|||||  
Db 212230 CTTTCATCAAGTTTGGCTTGTGCCCTGTCCCTGTGCTAATTAATCTTATTTATGCAA 212171  
QY 301 AGCTAAGTAATATTTTAAAGTACAGAGCTTATTTATGCAAGTGGAAAGCCCTTTTAA 360  
|||||  
Db 212170 AGCTAAGTAATATTTTAAAGTACAGAGCTTATTTATGCAAGTGGAAAGCCCTTTTAA 212111  
QY 361 TTTTCTTTGCCCCCTGTCGCCACTGTAATTTTAVCCGCTACGCGATGTTTACATCCTACAG 420  
|||||  
Db 212110 TTTTCTTTGCCCCCTGTCGCCACTGTAATTTTAVCCGCTACGCGATGTTTACATCCTACAG 212051  
QY 421 AATTTGCTGACCGTTTACAGGCAATCCCTACCTCCAGAGATTGCTAGACTGCTTGCATCT 480  
|||||  
Db 212050 AATTTGCTGACCGTTTACAGGCAATCCCTACCTCCAGAGATTGCTAGACTGCTTGCATCT 211991  
QY 481 TAAAGAACTGACAAATTTGCTGCATTTATGTACTTGTGCTGAACATATGGGAAAGCCTCATGC 540  
|||||  
Db 211990 TAAAGAACTGACAAATTTGCTGCATTTATGTACTTGTGCTGAACATATGGGAAAGCCTCATGC 211931  
QY 541 TATCTCTAATGTTTCTGGGGAATTTGCTAATGAATTAACAAAATCCAGAAAGCAAGCCTT 600  
|||||  
Db 211930 TATCTCTAATGTTTCTGGGGAATTTGCTAATGAATTAACAAAATCCAGAAAGCAAGCCTT 211871  
QY 601 TCTACGCTCTTTTGGATGATGCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATATG 660  
|||||  
Db 211870 TCTACGCTCTTTTGGATGATGCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATATG 211811  
QY 661 TTTGGGCTTCAAGTTTACAGACTTCGTTCTGGAAGGTATGATCTTGGGAAATTTCTT 720  
|||||  
Db 211810 TTTGGGCTTCAAGTTTACAGACTTCGTTCTGGAAGGTATGATCTTGGGAAATTTCTT 211751  
QY 721 TACCTCTTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
|||||  
Db 211750 TACCTCTTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 211691  
QY 781 GGATCAATAGAACGATATGACGATCTGCTTCTAATATCGAAGAAATGCAAAAG 840  
|||||  
Db 211690 GGATCAATAGAACGATATGACGATCTGCTTCTAATATCGAAGAAATGCAAAAG 211631  
QY 841 GGAAGAAAGGTCTAAGCTTAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 900  
|||||  
Db 211630 GGAAGAAAGGTCTAAGCTTAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 211571  
QY 901 CTCCTTATATCTTTTATTAAGCTCTTGGTATATGCTATGATTTGATTTGATTTAA 960  
|||||  
Db 211570 CTCCTTATATCTTTTATTAAGCTCTTGGTATATGCTATGATTTGATTTGATTTAA 211511  
QY 961 TCGAAGTACTTGGAAAAAGTACGTAAGCTGAATATCTAATATGATGATATGATG 1020  
|||||  
Db 211510 TCGAAGTACTTGGAAAAAGTACGTAAGCTGAATATCTAATATGATGATATGATG 211451  
QY 1021 AGTTCATGGGGAATCTCTCTTCTGAGCTGGCGATATCCGTATCATATGATATTTG 1080  
|||||  
Db 211450 AGTTCATGGGGAATCTCTCTTCTGAGCTGGCGATATCCGTATCATATGATATTTG 211391  
QY 1081 TTTGGTGAACGATCATTCGTAATATTTGATGATGATGATGATGATGATGATGATGATGAT 1140  
|||||  
Db 211390 TTTGGTGAACGATCATTCGTAATATTTGATGATGATGATGATGATGATGATGATGATGAT 211331  
QY 1141 TGGTTCTCTAACAAGGATAGCTTTTCTTGGCTCTTGTATCTTAAACCAAGCTTCTG 1200

```

|||||
Db 211330 TGGTCTCTACAGATGCTTTCTTCTGCTCTTGTATCTTTAGAACCAAGCTTCG 211271
OY 1201 GGCTGTGCTATGTCGTGTACACCTCTCTATGCTAGCTGTGGTTGCGAGCTATAC 1260
Db 211270 GGCTGTGCTATGTCGTGTACACCTCTCTATGCTAGCTGTGGTTGCGAGCTATAC 211211
OY 1261 AGAATATCTTTCGAATACCAAAATATGCTCTTTGATCACTAAAGAAATGGCCT 1320
Db 211210 AGAATATCTTTCGAATACCAAAATATGCTCTTTGATCACTAAAGAAATGGCCT 211151
OY 1321 AATATCCCTTTCGAACCAAGCAAAAGTCAAGAGTGTCTATTTGATGTAGTTGCCG 1380
Db 211150 AATATCCCTTTCGAACCAAGCAAAAGTCAAGAGTGTCTATTTGATGTAGTTGCCG 211091
OY 1381 CCGGCTTCGGAATACAGAGAGCTTTATTCACACAGAGTTGCTGCTATCTGCGAA 1440
Db 211090 CCGGCTTCGGAATACAGAGAGCTTTATTCACACAGAGTTGCTGCTATCTGCGAA 211031
OY 1441 GATTGAGCTATGACCCCTTATCTGAGTGATTCCTTTTCACTATGCTATTTGGT 1500
Db 211030 GATTGAGCTATGACCCCTTATCTGAGTGATTCCTTTTCACTATGCTATTTGGT 210971
OY 1501 TGGTTTCTCACTAAGATTAAACAAATATCTTTAGCGAGTCTGCTTTAAAGAACAG 1560
Db 210970 TGGTTTCTCACTAAGATTAAACAAATATCTTTAGCGAGTCTGCTTTAAAGAACAG 210911
OY 1561 AAGTGGCTCAAGAGATTAGCTCTGCTCTTCTTCAATAGAGTTGCTCTTACTCTTGT 1620
Db 210910 AAGTGGCTCAAGAGATTAGCTCTGCTCTTCTTCAATAGAGTTGCTCTTACTCTTGT 210851
OY 1621 GATCCCTACTGCTTTT 1637
Db 210850 GATCCCTACTGCTTTT 210834

```

RESULT 4  
AA201425  
ID AA201425 standard; DNA; 1038602 Bp.

```

XX AC AA201425;
XX DT 07-OCT-1999 (first entry)
XX DE Complete genome sequence of Chlamydia trachomatis.
XX KM Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX KM paratrachoma; inclusion conjunctivitis; genital disease; perithelitis;
XX KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX KM Bartholinitis; pneumonia; venereal lymphogranulomatosis; ss.
OS Chlamydia trachomatis.
XX PN W09928475-A2.
XX PD 10-JUN-1999.
XX PF 27-NOV-1998; 98WO-1B01939.
XX PR 04-NOV-1998; 98US-0107077.
XX PR 28-NOV-1997; 97FR-0015041.
XX PR 17-DEC-1997; 97FR-0016034.
XX PA (GEST ) GENSET.
XX PI Griffiths R.
XX DR WPI; 1999-371125/31.
XX PT Genome sequence of Chlamydia trachomatis
XX PS Claim 1; Page 373-656; 1755bp; English.
XX

```

CC The present sequence represents the complete genome of Chlamydia trachomatis. Open reading frames (ORFs) of the genome encode polypeptides AY36754-137949. The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conjunctivitis, genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perithelitis, Bartholinitis, CC pneumonia in breast feeding infants, and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.

Sequence 1038602 Bp; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 53.5%; Score 876; DB 20; Length 1038602;  
Best Local Similarity 73.0%; Pred. No. 18-229;  
Matches 1155; Conservative 0; Mismatches 421; Indels 7; Gaps 2;

```

OY 1 GAATATTAACATATCAGAAATAGAAATTAAGATTTTCAAGAGGT-AAATATGACAAA 59
Db 392356 GATCAACAAAAGAAACCTTAACATTAACCTTTTCAGAGGCTAAATATGACTCAA 392415
OY 60 ACCGAGAAAAACCTTTTGGAAAAATGCGCTCTTCTTGCGCGGATACATACTACAGAG 119
Db 392416 ACCGCGAAAAACCTTTTGGAAAAATGCGCTCTTCTTGCGCGGATACATACTACAG 392475
OY 120 CTAAAGAAAGTCTGCAATGTTCTCTAATGTTCTTGTATTAATCAATTAATACAGTG 179
Db 392476 CTGAAGAAAGTCTGCAATGTTCTCTAATGTTCTTGTATTAATCAATTAATACAG 392535
OY 180 TTACGGATACAAAACACTCTTATTTGGAGCTCTGTTCTGTGTGACAGGCAATA 239
Db 392536 TTGAGAGATACAAAACACTCTTATTCGTACAGACCGGAGCTGAGACAGGCAAT 392595
OY 240 CCTTCATCAAGCTTTTGGAGTGTGTGCTGCTGCTATATCTTATGCTATTTATTTGA 299
Db 392596 CCTTCATTAAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 392655
OY 300 AAGCTAAGTAATATTTTAAAGTAAGAGGCTCTTATTTATGACAGTGGGAAAGCCCTTTT 359
Db 392656 AAGCTAAGCAATATTTTGAAGAAACAGGCTCTTATTCGAGCTGCTACACATCGTT 392715
OY 360 ATTTCTTTGCGCTGTTCCCGACTTAATTTATCCGCTACGAGATTTTACATCTTA 419
Db 392716 GATTCTTTCGCACTGTTCCCTGCTGATCTACCTGCGCATATTTCTTCAACCCAMA 392775
OY 420 GAATTTGCTGACCGTTTACAGGCTATCTACCTCCAGATTTGCTAGACGCTGCTCATC 479
Db 392776 GCTTTGCTGATATCTTACAAATCATCTTCTCCGATTTATGAGATTCATTTGCGATG 392835
OY 480 TTAAGAACTGACATTTGCTGATTTTATGATCTGCTGCACTATGAGGAGAGCGTCATG 539
Db 392836 CTAGCGAAGCTGACATTTGCTGATTTTATGATCTGCTGCACTATGAGGAGAGCGTCATG 392895
OY 540 CTATCTCTAATGTTCTGAGGATTTGCTAATGAATTAACAAAATCCACAGCAAGCGT 599
Db 392896 CTCTCTTTATGTTCTGAGGATTTGCTAATGAATTAACAAAATCCACAGCAAGCGT 392955
OY 600 TTCTACGCTCTTTTGGATTTGAGAGCTATATTTTCTTCTAGCTCTGCTGCTGCTGCT 659
Db 392956 TTCTACGCTCTTTTGGATTTGAGAGCTATATTTTCTTCTAGCTCTGCTGCTGCTGCT 393015
OY 660 GTTTGGCTTCAAGTTGAGAGCTTCTCTTCTGAGAGTGTAGATCTTTGGGAAATTTCT 719
Db 393016 ATCTGCTTTCTAAATTTACGCTTACCTTTGGGAAAGGCTTGAACCAAGGAGTCAAC 393075
OY 720 TTACGCTTTTATGCTATGATATTTGATCTGCACTGTTCTTATGCGCCAGTACTG 779
Db 393076 CTCTATTTCTTAAGCTATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 393135
OY 780 TGGATCAATAAGAGCTATTTGACCGATCTGCTCTTATTAATCCAGAAAGCAATGCAAAAG 839
|||||

```

Dh 393136 TGGATGAACCGCTACGTCGTTACAGATTCCTACATCTCCAGAGCTCAAAGCT 393195  
OY 840 GGGAAAAAGGTGCTAAACCTAAATGAATATGAAGATAGCTTCCTATCTGTATAGA 899  
Dh 393196 AAGAAA-----TCTAAACCTAAGATGACATGGGGAAGAGCTTCCTATCTGTATAGA 393249  
OY 900 TCTCCTTATATCTTTTATTAACCTCTTGGTATGCTTATGCTTATGCTTATTAACCTTA 959  
Dh 393250 TCTCCTTATATCTTCTTCTACCTCTTCTAGTATTTGCTACAGGAATCGCATTTAAACCTC 393309  
OY 960 ATCGAAGAGCATTTGGAAGAAAGAGCTGCAATATCCCAATATGAATGATCATATATG 1019  
Dh 393310 GTTGAAGTACTGGAAGAAAGCACTCAAAATGCAATTCCTTAATCCAAATGCAATACAC 393369  
OY 1020 GAGTTCATGCGAAGCTTCTCTTCTGAGATGCGCTAGTATCCGCTACTTATCATGCTATTT 1079  
Dh 393370 GCATTTATGAGGAGACCTCTCTTCTGAGAGAGTGTATCTGATATGCTATGCTCTTC 393429  
OY 1080 GTTGGTGTACAGCTATTCGTAATTTGGATGCTTAACCTGAGACCCCTACTCTCTCTTC 1139  
Dh 393430 ATCGGTGGTAACTGATATCCGTGATTCGCTGTTAACTGAGAGCTTTGCTTACCTTATC 393489  
OY 1140 ATGGTTCCTTACAGATGATGCTTTCTGCTGCTTGTATGTTAAAGCAACGCTCTCT 1199  
Dh 393490 ATGCTTTCGTAACAGAGAGCTCTCTTCTGCTGCTTATCTTCAAGATCATGCTTACA 393549  
OY 1200 GGGCTGTCGCTATGTTGGGTACAACCTCTCATGCTAGCTGTGTTGCTCGAGCTATA 1259  
Dh 393550 GGTATTGTTGCTGCTTGGGGAACACACCACTAATGCTAGCTGTTGTTGCTCGAGCCTAT 393609  
OY 1260 CAGATATCTCTTGAATCCCAAAATAGCTCTTGTGACTCAACTTAAGAAATGGCC 1319  
Dh 393610 CAAATATCTCTTCAAAATCACTAATATGCTCTTGTGAAGCAACTTAAGAGCTGGCT 393669  
OY 1320 TATATCCCTCTTGAACAGAGCAAAAGCAAGCTAAGCTGATGATGATGATGCTGCC 1379  
Dh 393670 TACATCCCATTCGATCAAGAGCAAAAGCTTAAGGAAAGCTGATGATGATGCTGCC 393729  
OY 1380 GCCCGCTTCGGAATATCAAGAGAGCTTATTCACAACAAGTTCCTGCTATGCTGGA 1439  
Dh 393730 GCTGATATCGTAATATGCGGAGATCTTGTGATCCAAAGAGCTGCTGCTTGGCGGA 393789  
OY 1440 AGTATGAGCATGATACCCCTATATGTCAGATTCCTCTTATGATCATGCTCTTTGG 1499  
Dh 393790 AGCATCGCGGCTATGATCCATCTCTAGCTGCTGCTCTCTTGAATATATCATGCTGCG 393849  
OY 1500 TTGCTTTCTGCACTAAGTTAAACAACCTATCTTAGCGCAGTCTGCTTAAAGACAA 1559  
Dh 393850 TTGACATCTGCACTAAGTTAAACAACCTGTTCTTGGCTGCTGCTGCTTAAAGACAG 393909  
OY 1560 GAGTGGCTCAAGAAATTCAGC 1582  
Dh 393910 GAATTAGNAGAAAGCTGCAGCAGC 393932

RESULT 5  
AAF61374  
ID AAF61374 standard; cDNA: 1896 BP.  
XX  
AC AAF61374:  
XX  
DT 25-MAY-2001 (first entry)  
XX  
Dh Potato chloroplast ATP/ADP translocator cDNA.  
XX  
DE ATP/ADP translocator: chloroplast; plant; adenosine triphosphate;  
KW adenosine diphosphate; animal feed; energy; biosynthesis; potato;  
KW nutritional value; ss.  
XX  
OS Solanum tuberosum.  
XX  
PN WO200120009-A1.

PD 22-MAR-2001.  
XX  
XX  
PF 05-AUG-2000; 2000WO-EP07625.  
XX  
PR 15-SEP-1999; 99DE-1044212.  
XX  
XX  
PA (BADI ) BASF AG.  
PI Reinold A, Geigenberger PL, Neuhaus H, Graeve-Kampfenkel K;  
PI Moellmann T, Tjeden J;  
XX  
XX WPI: 2001-244814/25.  
XX  
XX  
PT Transformed plant with an altered content of essential amino acids and  
PT having improved nutritional quality, for use as an animal feed,  
PT comprises a modification in the ATP/ADP translocator gene -  
PS Example 4; Fig 2; 32pp; German.  
XX  
XX  
CC This invention describes a novel transformed plant (A), and its  
CC descendants, in which the regulatory sequences and/or the gene copy  
CC number of an ATP/ADP (adenosine tri-/di-phosphate) translocator gene  
CC (I) has been altered so that, compared with the wild-type plant, it  
CC has a different content of one or more amino acids (aa). The invention  
CC also describes (1) (I), for use in plants, having an approximately 1.7  
CC kb sequence (S1), given in the specification, from Arabidopsis thaliana,  
CC EMBL 249227; (2) a genetic construct (GC) containing the (1) and operably  
CC linked regulatory sequences; (3) a vector containing the (1) or GC; and  
CC (4) seeds, tissues, cells and replicative materials of (A). (A) are  
CC useful for human or animal feeding, also (including their cells, tissues  
CC and extracts) in agriculture, the animal feed and pharmaceutical  
CC industries, and the health service. Altering the activity of (1)  
CC increases the amount of ATP in chloroplasts and the amount of energy  
CC available for biosynthesis. Specifically (A) have increased contents of  
CC one or more essential aa, so are of greater nutritional value than  
CC wild-type plants.  
XX  
XX  
SQ Sequence 1896 BP: 473 A; 369 C; 445 G; 609 T; 0 other:  
Query Match 23.2%; Score 380.2; DB 22; Length 1896;  
Best Local Similarity 55.1%; Pred. No. 2.6e-94;  
Matches 804; Conservative 0; Mismatches 628; Indels 27; Gaps 2;  
OY 74 TTTTGGAAATTTGCGCTCTTCTGCTGTCGCGCATACATCAACAGCTTAAGAAATGCTCT 133  
Dh 285 TGTGGAAGAAAGGACCACTAAGTTATGAGGATGAACCTTGTGACCTTAAAGAAATATAT 344  
OY 134 GCCATGTTCTTAATGTTCTCTGATTAACATTAACATATACGTTGTAACCGCATACAA 193  
Dh 345 ACCACTGGGCGATGTTCTTTGATTCGTGTTAATTAATCAATCCTTAGGATATCA 404  
OY 194 AGACACTCTTATTTGGAGCTCCTGTTGTCGTCAGAGCAATACCTTTCATCAAGTT 253  
Dh 405 GGAATGTTGTTGTTAAGCTAAGCTAAGAGGTCAGTGTGAGATTTATCCCTTTCTGGAAC 464  
OY 254 TTGCTTGTGTCCTGCTGCTATTAATCTTATGCTTAATTAAGCAAGCTAAGTATAT 313  
Dh 465 TTGGGTGAATTTGCTATGCTATGATGATGATGATGATGATGATGATGATGATGATG 524  
OY 314 TTTAAGTAAAGAGGCTTATTTATGAGTGGGAAGCGCTTTTAATTTCTTGGCCCT 373  
Dh 525 GTTCCGAGCTTAATTTATCGCTACGAGATGTTTATCATCTACAGAAATTTGCTGACCG 584  
OY 374 GTTCCGAGCTTAATTTATCGCTACGAGATGTTTATCATCTACAGAAATTTGCTGACCG 433  
Dh 585 GTTGGTTTGTGTTTGTATCTCTTAGCAATTAATTTACCTTACAGCTTTTGTGATTA 644  
OY 434 TTTAAGGCGCATCTACCTCAGATGCTAGAGCTGTTGCCATTTAAGAAATGAGAC 493  
Dh 645 GCTTCTCAATACCTGCTGTCAGAAATTTCTTGACCAATTCATCTATGAGAGATCTGGAG 704  
OY 494 ATTGCTGATTTTATGATGCTGATGATGAGGAAAGCTCATGCTATGATATGTT 553

RESULT 6  
AAC36046

PR	18-JUN-1999;	99US-0139459
PR	18-JUN-1999;	99US-0139460







D	b	620	AAACTTTCCAAATGTTCTCTCCAAAAGGCGCTCTTTAAACGTGTATGTTCCTTTCAAT	673
Q	y	360	ATTTCCTTTGGCCCTGTTCCCGACGTAAATTTATCCGCTACGGAGTGTTTACATCCTACA	419
D	b	680	GTTCTACTTTGGAGCCTTTGGTTCTTGAGATGACCCCTCAGCAATTTGATTCATCTCGAA	739
Q	y	420	GAATTTCCACCGCTTTACAGGCCATCTACCTCCAGATTTGCTAGAGACTGCTTCCATC	479
D	b	740	GCTCTTCTGATTAAGCTCTTTGGCAACACCTCGGCCCGCCAGTTTCATGGGTCTCTCCCAATC	799
Q	y	480	TTAAGAAACGACATTTGCTGCATTTTATATGATCTTTCGACACTATGGGGAAGCCTCATG	539
D	b	800	ATGAGGATTTTGGAGTTTCTGTGTTCTATGATCATGCTGACACTTTGGGGATGTTGTG	859
Q	y	540	CTATCTCTAATGTTCTGGGGATTTGCTAAATGAATTTACAAAAATCCAGCAAGCAAAAGCT	599
D	b	860	GTTTCAGTCTCTCTGCTGGGATTTTGGCAACCGAATTTACAACTGTTGACGAAACCCAAAAG	919
Q	y	600	TTTACGCTCTTTTCGCTATCGGAGCTAAATATTTCTTAAGCTTTGCTGTCGAAT	659
D	b	920	TTTCTATCTCTGTTTGGACTTTGGGGCAATGTGTCACTATCTTCACAGAAAGCACTGTG	979
Q	y	660	GTTTGGGCTTCAAGTGTGACGCTCCGTTTTCGAAAGGTGTGATATCTTTGGGGAATTTCT	719
D	b	980	AAATATTTTCTCTAATATGAGAAAGAAATCTTGGTCTGGAGTTGATGGCTGGGCTTTTCA	1039
Q	y	720	TTACGCTTTTGTGATGGCTATGACATATTTGTATCTGCACTGTTCTTATGCGCAGTTACTGG	779
D	b	1040	TTAAACCTTGTGATGATGATATTTGTGCGGGGATGGCTCGGCATCTGTTTCTCTACTGG	1099
Q	y	780	TGATCTAATTAAGAACGTATTGACCGATCTCGCTCTTATATCCAGAAAGAAATGCAAAAG	839
D	b	1100	TGGGTGAAT-----GATATGTGCCCCCTCCCAACCGGTAGC	1139
Q	y	840	GGGAAAAAGGTGCTAACTTAATGAAATGAAATGAAAGATGAGTCTCTCATCTGTATAGA	899
D	b	1136	AAGAGAGAAAGGTGAACACCGATGGGACATGAGAGGCTTGAAAGTCTTGGTGTC	1199
Q	y	900	TCTCTTATATTCCTTTTATTAACCTCTCTTGTTGTTATTTGCTATGTAATTTGCAATTA	959
D	b	1196	TCACCATACATTTAGGAGATCTTGCTACTTTGGTGGTTCGATATGAAATAGTATCAACT	1255
Q	y	960	ATCGAGTGACTTGGAAAAGCTCAGCTGGAACCTGCAATATCTTAATATGATGACTATAGT	1019
D	b	1256	GTTTGAAGTCAATGAAATCAAAAGCTTAAAGCTCAATCCCTAGCCGCAAGCAATATCA	1315
Q	y	1020	GAGTTCATGGGGAACCTCTCTCTGAGCTGGGAGTATCCGTACTTATCATCTATTT	1079
D	b	1316	GCATTTATGGGCGCTTCAACCTTCCACAGGATTTGCAAC--ATTCACAATGATGGTT	1372
Q	y	1080	GTTTGTGTGAAGCTATTCCTAAATTTTGGATGGTTTAACGAGCCCTGATGCTACTGTG	1139
D	b	1373	CTAAGCCAAATACGTGTTTAAAGAGTATGGTTGGGGAGTACGTGCAAAAGTATCACCAAC	1432
Q	y	1140	ATGCTTCTCTAACAGGATATCGTTTCTTCGCTTGTATTCCTTTAGAAACCAAGCTCT	1199
D	b	1433	GTTCTCTATTAAGCCGCTGTGCTTCTCTCTCTGATCTGTTTGGTGGCCATTTGCA	1492
Q	y	1200	GGGTGTGCTATGTTCCGTACAACTCTCTCATCTACTAGCTGTGTGTTGTGCGAGCTTA	1258
D	b	1493	CCATTTGTTCCCAAGCTGTGATGAACACCGCTACCTGCGAGGTGTAGCTGTGTGGTCCCTC	1532
Q	y	1260	CAGAAATTTCTTTCGAAATCCACAATAATAGGCTCTTTGATCTCAACTTAAGCAAAATGGCC	1319
D	b	1553	CAGAAATTTCTTCAGCAAGAGTGGCAAGTACAGTTGTGTGATCTCTTGCAGAAAGAAATGGCT	1612
Q	y	1320	TATATCCCTCTTGACCAAGCAAAAGTCAAGAGTAAAGCTCTATTGATGTAGTTGCC	1379
D	b	1613	TATATCCATTTGATGAGGACACCAAGGTTTAAAGGCAAAAGCTGCAATGATGTGCTGCG	1672
Q	y	1380	GCCGCTTCGGAATATCAGAGAGCTTTATATCAACAAAGTTTGCTGTTATCTGTGGA	1439
D	b	1673	AACCAATTTGGGGAATTCAGCGCGTGTCTTAATACACAGCTTATCATCTTTCATCTGCGC	1732

QY	1440	ACTATTGGAGCGATGAGCCCTTATCTGACGATGATCTCTTTTCATCTGCTATTGG	1499
DB	1733	TCAGTCGCCAATTCACACACTTACCTTGAGTCATCTGCTGATTAAGTCACTGCATGG	1792
QY	1500	TTGGTTTCTGCACAA	1512
DB	1793	TTAGCAGCAGCTTA	1805
RESULT 7			
AAFE1373			
ID	AAFE1373	standard; cDNA; 1770 BP.	
XX	AAFE1373:		
AC			
XX			
DT	25-MAY-2001	(first entry)	
XX			
DE	A. thaliana chloroplast ATP/ADP translocator cDNA.		
XX			
KM	ATP/ADP translocator; chloroplast; plant; adenosine triphosphate;		
KW	adenosine diphosphate; animal feed; energy; biosynthesis;		
KM	nutritional value; ss.		
XX			
OS	Arabidopsis thaliana.		
XX			
PM	WO200120009-A1.		
XX			
PD	22-MAR-2001.		
XX			
XX	05-AUG-2000; 2000MO-EP07625.		
PF			
XX	15-SEP-1999; 99DE-1044212.		
PR			
XX			
PA	(BADI ) BASF AG.		
XX			
PI	Reinold A, Geigenberger PL, Neuhaus H, Graeae-Kampfenkel K;		
XX	Moehlmann T, Tjaden J;		
DR	WPI: 2001-244814/25.		
XX			
PT	Transformed plant with an altered content of essential amino acids and		
PT	having improved nutritional quality, for use as an animal feed,		
PT	comprises a modification in the ATP/ADP translocator gene		
XX			
PS	Claim 6; Fig 1; 32pp; German.		
XX			
CC	This invention describes a novel transformed plant (A), and its		
CC	descendants, in which the regulatory sequences and/or the gene copy		
CC	number of an ATP/ADP (adenosine tri-diphosphate) translocator gene		
CC	(I) has been altered so that, compared with the wild-type plant, it		
CC	has a different content of one or more amino acids (aa). The invention		
CC	also describes (1) (I), for use in plants, having an approximately 1.7		
CC	kb sequence (S1), given in the specification, from Arabidopsis thaliana,		
CC	EMBL Z49227; (2) a genetic construct (GC) containing the (I) and operably		
CC	linked regulatory sequences; (3) a vector containing the (I) or GC; and		
CC	(4) seeds, tissues, cells and replicative materials of (A). (A) are		
CC	useful for human or animal feeding, also (including their cells, tissues		
CC	and extracts) in agriculture, the animal feed and pharmaceutical		
CC	industries, and the health service. Altering the activity of (I)		
CC	increases the amount of ATP in chloroplasts and the amount of energy of		
CC	ATP available for biosynthesis. Specifically (A) have increased contents of		
CC	one or more essential aa, so are of greater nutritional value than		
CC	wild-type plants.		
XX			
SO	Sequence 1770 BP; 441 A; 383 C; 431 G; 515 T; 0 other;		
Query Match	20.2%; Score 330; DB 22; Length 1770;		
Best Local Similarity	54.5%; Pred. No. 1.7e-80;		
Matches 769; Conservative	0; Mismatches 610; Indels 33; Gaps 4;		
QY	121	TAAAGAAAGTTTGGCAATGTTCTCAATGTTCTCTGATTAACATTTAACTTAAGTGT	180

Db 335 TGAAGAGATTATCCCTTTAGCATGATGTTCTTTGATTCTTTTCATTAACACAAATTC 394  
 Qy 181 TACGAGATACAAAGACACTCTATTGTGGAGAGCTCTGTGTTGTGTGACAGAGCAATAC 240  
 Db 395 TGAGGAGATACAAAGAGATCTCTGGTGGACGCGAAGGAAGTTCTCTGATGATTATAC 454  
 Qy 241 CTTCATCAAGTTTGGCTGTGTGCTCCCTGTGCTATTTATCTTAAGCTTATTTATGCA 300  
 Db 455 CTTCCTTGAAACACTGGGTGAATCCTCTATGGCATTTGCTTATGCTCCCTACACTA 514  
 Qy 301 AGCTAAGTATATTTTAAGTAAAGCAGCTTATTTATGACGGGAGAGCCCTTTTAA 360  
 Db 515 AACCTCCAAATGTTCTCCACAGAGAGCTGTGTTTACAGCTTATTTGCTCCCTTCATCA 574  
 Qy 361 TTTTCTTTGGCCCTGTCCCGACTGTAATTTATCCGCTACGAGATGTTTATACATCTACAG 420  
 Db 575 TCTACTTTGGGGGCTTTGGTTTGGTCATGATCCCTCAGACACTATTTATACCCGGAG 634  
 Qy 421 AATTTGCTGACCGTTTACAGGCAATCTACCTCCAGAGATTGCTAGAGACTGCTGCATCT 480  
 Db 635 CTCTCGGAGATAGCTCTCTTACAACCCCGGCGCAAGATTCATGGCTCTATTGCAATAT 694  
 Qy 481 TAAGAACTGGACATTTGCTGATTTTATGACTTGTGCACTATGGGAGCCGTCATGC 540  
 Db 695 TGGGATTTTGGAGTTTCTGTTTGTATGTTATGCTGAGCTTTGGGTAGTGCGTGG 754  
 Qy 541 TATCTTAATGTTGCGGAGATTGCTAATGAATTAACAATAATCACAGACCAAGAGCTT 600  
 Db 755 TCTGAGTTCTCTTCTGGGGCTTTGCTATATCAGATCACAACACTGTGATGACCAAGAAAT 814  
 Qy 601 TCTACGCTCTTTGCGATCGAGCTAATATTTCTTAAGCTTGTGCTGCTGCAATG 660  
 Db 815 TCTATCTTGTGTGGCATGTGAGCCAAATGTTGCACTGATTTTCTCAGGAAGAACCGTGA 874  
 Qy 661 TTTGGGCTCAAGTGTGAGCTTCCGTTCTGAAGGTGTGATCCTTGGGGAATTTCTT 720  
 Db 875 AATACCTTCTTACCTTGGAAGAATCTTGCTCCTGGAGTTGA-----CGCAGATTTCTG 929  
 Qy 721 TACCTCTTTGATGCTATGACTATTTGATGCTGACTGTGTTATGCGCAGTTACTGAT 780  
 Db 930 TGAAGGCAATGATGAGCATTTGTGTGGGAATGGAGCTG--CATTTGTCTCTCATATGCT 987  
 Qy 781 GGATCAATAGAACGTATGACCGATCTCCCTCTTATATCCAGAGAAGAAATGCAAAAG 840  
 Db 988 GGGTCGATAGA-----TATGTTCTCTTCCAAACCCGTAGCA 1024  
 Qy 841 GGAAGAAAGGTCTAAACCTAAATGAATATGAAGAGCTCTCTCATATCTGATAGAT 900  
 Db 1025 AGAACAAGAGAGAGAACCCGAGATGGAGACGATGGAAGCTTGAAGTTCTTGGATCAT 1084  
 Qy 901 CTCCTTATATCTTTTATTAACCTCTGTGATTATGGCTATGATTTGCAATTAACTTAA 960  
 Db 1085 CACCATCATATTAGAGATCTTCTACTTTAGTGTGGCATAGGATATTAGTATCATATCTG 1144  
 Qy 961 TCGAAGTACTTGGAAAAGTCACTGAAATCTCAATATCTTAATATGATGACTATAGTG 1020  
 Db 1145 TGAAGATCACATGGAATCAAAAGCTTAAAGCTCAGTCCAGCCGCAATGATCACTAG 1204  
 Qy 1021 AGTTCATGGGGAACCTCTCTGAGCTGGCGATGATATCCGTAATCATATGCTATTTG 1080  
 Db 1262 TCAAGCAATACGATTTCAATATAGTATGTTGGGAGTAGTACGCAAGATCAACCCCACTG 1321  
 Qy 1081 TTTGGTGAACGTCATTCGTAATTTGGATGTTAACTGAGACCCCTAGTCACTCTCTGCA 1140  
 Db 1262 TCAAGCAATACGATTTCAATATAGTATGTTGGGAGTAGTACGCAAGATCAACCCCACTG 1321  
 Qy 1141 TGGTTCTCTTAACAGGATGCTTTCTGCTGTTATCTTATCTTAAGAAACCAAGCTTCTG 1200  
 Db 1322 TTTGCTATTGAGATGCTGTTGCTGTTCTCTTAATATTTTGGGGCCCATTTGCGAC 1381  
 Qy 1201 GGGGCTGCTGATGTTGGGTCAACTGCTCTCACTGATGCTGTGTTCTGCGACTATAC 1260  
 Db 1382 CACTTGTTCGAAGCTGTGATGACACCGGCTACTTGCAGCTGTGATGCTGCTGCCCTTC 1441

Qy 1261 AGAATATCTTTCGAAATCCAAATATACGCTCTTGTGACTCAACTTAAGAAATGGCCT 1320  
 Db 1442 AGAATATCTTTCGAAATCCAAATATACGCTCTTGTGACTCAACTTAAGAAATGGCCT 1501  
 Qy 1321 ATATCCCTCTTGACCAAGACAAAGATCAAGATGAGCTGCTATGATGATGATGCGG 1380  
 Db 1502 ATATCCCTCTTGACCAAGACAAAGATCAAGATGAGCTGCTATGATGATGATGCGG 1561  
 Qy 1381 CCCGCTTGGAAATTCAGAGAGAGCTTAAATCCAAAGATGAGCTGCTATGATGATGCGG 1440  
 Db 1562 ACCCATTAAGAAATTCAGAGAGAGCTTAAATCCAAAGATGAGCTGCTATGATGATGCGG 1621  
 Qy 1441 GTATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
 Db 1622 CACTAGCAATTCAAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1681  
 Qy 1501 TGGTTTCTGCACTAAGTTAAACAACATATTC 1532  
 Db 1682 TAGCTGAGCTAAGTCTGCTGAGGAGACAGTTC 1713  
 RESULT 8  
 AAF84486  
 ID AAF84486 standard; DNA; 1823 BP.  
 XX  
 AC AAF84486:  
 XX  
 DT 02-JUL-2001 (first entry)  
 XX  
 DE Chlamydia pneumoniae Npt2cp (ADP/ATP translocase) gene.  
 XX  
 KW Npt2cp: ATP translocase; respiratory tract infection; pneumonia;  
 KM bronchitis; sinusitis; atherosclerosis; asthma; vaccine; antimicrobial;  
 KW antibacterial; diagnosis; ds.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 101..1723  
 FT /tag= a  
 FT /product= "Chlamydia pneumoniae Npt2cp protein"  
 FT /function= "ADP/ATP translocase"  
 XX  
 PN W0200121803-A1.  
 XX  
 PD 29-MAR-2001.  
 XX  
 PF 15-SEP-2000; 2000WO-CA01087.  
 XX  
 PR 17-SEP-1999; 99US-0154326.  
 XX  
 PA (AVET ) AVENTIS PASTEUR LTD.  
 XX  
 PI Murdin AD, Oomen RP, Wang J, Dunn P;  
 XX  
 DR WPI: 2001-316102/33.  
 DR P-PSDB: AAB73452.  
 XX  
 PT New Npt2cp (ADP/ATP translocase) polypeptides and nucleic acids  
 PT encoding the polypeptides useful for treating, preventing or diagnosing  
 PT Chlamydia infections, particularly infections caused by Chlamydia  
 PT pneumoniae  
 PS Claim 2; Fig 1: 79pp; English.  
 XX  
 CC This sequence represents the Chlamydia pneumoniae Npt2cp (ADP/ATP  
 CC translocase) gene. Chlamydia pneumoniae Npt2cp nucleotides and  
 CC proteins may be used for preventing, treating or diagnosing  
 CC chlamydial infections, particularly those caused by Chlamydia  
 CC pneumoniae. Chlamydia pneumoniae can be responsible for both upper and  
 CC lower respiratory tract infections. It is the third most common cause of  
 CC community acquired pneumonia, and can also cause bronchitis and

CC sinusitis. It is also linked to diseases other than respiratory tract  
 CC infections, being associated with atherosclerosis and asthma.  
 CC Npt2cp nucleotides may be used for the recombinant production of the  
 CC protein, as genetic vaccines, in the construction of vaccine vectors  
 CC such as poxviruses, and in the construction of attenuated Chlamydia  
 CC strains which can over-express an Npt2cp nucleotide, or which can  
 CC express it in a non-toxic, mutated form. Npt2cp polypeptides can be used  
 CC in vaccine compositions, and may also be used as diagnostic reagents for  
 CC detecting the presence of anti-Chlamydia antibodies in a sample.  
 XX

SO Sequence 1823 Bp; 450 A; 359 C; 396 G; 618 T; 0 other:

Query Match 18.4%; Score 301.4; DB 22; Length 1823;

Best Local Similarity 50.3%; Pred. No. 1.3e-72;

Matches 814; Conservative 0; Mismatches 776; Indels 27; Gaps 2;

OY 2 AATATAAACTATCAGATATGAAATATAAGTATTCAGAGGGTAAATATGACAAAAC 61  
 DB 52 AACTAAATGCTATAGCTTTATACGCCATTAAGTTGGAGAGGTTTATGACATGATC 111  
 OY 62 CGAATAAAACCTTTTGGAAATTCGGCTCTTCTGGCCGATATACATACAGAGCT 121  
 DB 112 AGAAGTGAACCTTTTGAAGGCTCGGGCATATCTTGTCTATTATATAATCAGAAAT 171  
 OY 122 AAGAAGATCTGCAATGTTCCATATGTTCTCTGATATACATTTAACTATACGGTGT 181  
 DB 172 TTCTAGATTGTTCCATATTTCTACTAGCGTTTTCGTGGCTTTAACTACTGCTGCT 231  
 OY 182 ACGCATACAAAAGACTCTTATTTGTGGAGCTCCTGTTCTGGTGCAGAGCAATACC 241  
 DB 232 GAAAAACATGAAGAATCTGCTGCTATGTCGTTTCAGATGCTGGGGGAGAGATATCC 291  
 OY 242 TTTTCATCATGTTTGGCTTTGTCGCCCTGTCCTATTAATCTTATATGCTTATATGCAA 301  
 DB 292 CTTCCTTAGGTTTGGGAATTTGCCGGAGCTGTATTTGTTACTATGTTATGGGTG 351  
 OY 302 GCTAAGTATATTTTAAAGTAAAGCAGCCTTATTTATGAGTGGAGAGCCCTTTTAAAT 361  
 DB 352 GTTAGGAGTGGCTATCTCGGGATATCCGCTTTTATTTATGCTTCATGCGCATTTCTGG 411  
 OY 362 TTTCTTGGCCCTGTCGCCGACTGTATTTATCCGCTACGCGATGTTTATACATCCACAGA 421  
 DB 412 TTTTTCCTTCCGTTTGGCTGATCATTTATCCGTAGGAGATACCTGATCTCAACGC 471  
 OY 422 ATTTCGTAACCTTTACAGGCCATCTCACTACAGAGATTGCTAGAGATCGTTGCCATCT 481  
 DB 472 TCTGCTATATAATTAACAAGAGCTCTTCCATCAAGAGACTCGTGTATTTATGATAGGT 531  
 OY 482 AAGAAGTGAACATTTGCTGATTTTATGTAATGTAATGTAATGTAATGTAATGTAATGTA 541  
 DB 532 CCGTACTAGGAGTATACAGTATTTATAGTATGTCAGAGCTGTGAGAGTGTGGTGTCT 591  
 OY 542 ATCTCTAATGTTTCGGGATTTGCTAATGAATTAACAAAATTCACAGCAAGACGTTT 601  
 DB 592 TTGCTATGTTTTCGGGACTATGCAATCAGATTAATTAATTAATTAATTAATTAATTA 651  
 OY 602 CTACGCTCTTTTTCGGTATGCGAGCTAATATTTCTTACAGTCTGCTGCGATATGCT 661  
 DB 652 TTACGCTCTTTATCAATACAGATTAATATCTCTCTCAATATGCGAGGAGAAATCTCCA 711  
 OY 662 TTGGCTTCAAAGTTGAGAGCTTCGTTTTCGAAGGTGATACCTTGGGGAATTTCTTT 721  
 DB 712 TTGGATGGGGAACAACATTTGTTGCGTACCTTTTGCATGTATGTTCTGCGACTGTGT 771  
 OY 722 ACGTCTTTGATGCTATGACTATTTGATCTGGAATTTGTTCTTAATGCGCAGTTACTGCTG 781  
 DB 772 AATGCTCAACTTGACCATGCTGATCACTGTTCTGTTTAAATATGATCTGCTATATAG 831  
 OY 782 GATCAATTAAGAAGCTATTTGACGATCTGCTGCTC-----TA 817  
 DB 832 GCGGATTAATCACTTTGACTATTGATAGCTTCATCCCTCCTCCTACATGACGCTGTTGGCAGA 891  
 OY 818 TAATTCAGAGAATGCAAAAAGGGGAAAAAGGTGCTAAACCTAATAATGATATGAAGA 877

DB 892 ACAGGAGACGACTACTGCTATATCTAAAGCAAAAAGAAAACCTAAAGCAAGCTAGAAA 951  
 OY 878 TAGCTTCCCTATCTTGTATAGATTCCTTATATTTCTTTATTAATCTCTGTTATTCG 937  
 DB 952 CCTTTTCTTACACCTCATCTACGTCGTATTTATTTAGGCGCTGGCTATATTTGCTATTC 1011  
 OY 938 CTAATGATTTTTCATTAATCTTAATGAGTGCATGGAAATCAAGTACAGCTGAATCTCAATA 997  
 DB 1012 CTAATATTTGTTGATTCATCTATTCGAAGTCTTTGGGAAGATCAAGATAGCCAGATTTTA 1071  
 OY 998 TCCATATATGAATGACTATATGATGATTCATGAGGGAACCTCTCTCTGAGCTGGCTAGT 1057  
 DB 1072 CAGTTCTCACAGATTAATTAATGAGTATATGATGATTAATCACTACCTCATATGGCTGCT 1131  
 OY 1058 ATCCGACTTATATATGCTATTTGTTGGTGTATACGCTCATTCGTAAATTTGATGTTAAC 1117  
 DB 1132 TTCTGATTAATGACAGCTGCTCTTACCGGACAGTGTATCCGTAATATGGGATGGACTGT 1191  
 OY 1118 TGGAGCCCTAGTACCTCCGCTATGATGTTCTCTACAGTATGCTTTCTTCTGCTTGT 1177  
 DB 1192 CGGTGCTTTAGTACTCCATTTGATATGTTAGTTTACAGAGCTGCTTTTTCGAACTAT 1251  
 OY 1178 TATCTTTAGAAACCAAGCT---TCTGGGCTGCTCCCTATGTTCCGTTACACTCTCTCAT 1234  
 DB 1252 TTTTGTGCAAAAAGACATCTATTTTGGGGAGTCTTGGATATGACACTCTGCG 1311  
 OY 1235 GCTAGCTGTGTTGTCGAGCTATACAGATATTTCTTTGAAATCCACAAATATACGCTCT 1294  
 DB 1312 TCTACTGCTGCTGAGCTGAGGAGTCAAAATGCTCTATCCCGGGAGCAAAATTTTACGTT 1371  
 OY 1295 CTTTACGCTCACTAAAGAAATGGCTATATCCCTCTTGACCAAGAGCAAAAGTCAAAAG 1354  
 DB 1372 CTTTATCAACCAAGAAATGGCTTTATCCACTCTCTCCAGAGGATTAATAATCATG 1431  
 OY 1355 TAAGCTGCTATTTGATATGATTTGTCGCCGCTTCCGAAATACAGAGAGCTTTTATATCA 1414  
 DB 1432 GAAACCCGGATTTGATGCTGCTTTCAAGATATAGAAAGTCTGGAGGCTCTTATATTTA 1491  
 OY 1415 ACAAGTTTGTCTGTTATCTGATGGAAGATTTGAGAGCTATGACCCCTTATCTTGCAGAT 1474  
 DB 1492 CCAAGGGGCTGTTGATTTTCTCTTCTGTTGTCAGCAAGTTTAAAGTCATGCGCTAGT 1551  
 OY 1475 TCTCTTATTAATGCTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1534  
 DB 1552 TCTCTCATTAATTAATGCTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1611  
 OY 1535 AGCGAGTCTGCTTTAAAGCAAGAAGTGCCTCAAGAAGATTCAGCTCGCTTC 1591  
 DB 1612 CTTAGAGCTGCTGATGCTGTAGCAACCTTGAACCAACCTTAAGAACCTTCTCTTTC 1668

RESULT 9  
 ABL74862  
 ID ABL74862 standard; cDNA; 284 Bp.  
 XX  
 AC ABL74862;  
 DT 14-MAY-2002 (first entry)  
 DE  
 XX Corn tassal-derived polynucleotide (cdps) spq ID NO:4236.  
 XX  
 KW Corn; corn tassal-derived polynucleotide; cdps; hybrid breeding; CDPS;  
 KW inheritance; characterstic; growth; development; disease resistance;  
 KW environmental adaptability; quality; yield; molecular marker;  
 KW multigene trait; plant breeding; corn tassal; gene; ss.  
 OS  
 XX Zea mays.  
 XX  
 XX US2001051335-A1.  
 XX  
 XX 13-DEC-2001.  
 PD  
 XX

PF 16-APR-1999; 9905-0294093.  
XX  
PR 21-APR-1998; 98US-082567P.  
XX  
PA (LALG/) LALGUDI R V.  
PA (ITOL/) ITO L Y.  
PA (SHER/) SHERMAN B K.  
XX  
PI Laljudi RV, Ito LY, Sherman BK;  
XX  
DR WPI; 2002-163647/21.  
XX  
PT Novel purified corn tassel-derived polynucleotide useful for  
PT determining altered gene expression, to recover regulatory elements and  
PT to follow inheritance of desirable characteristics through hybrid  
PT breeding programs  
XX  
PS Claim 1; SEQ ID 4236; 201pp; English.  
XX  
CC The present sequence describes a purified corn tassel-derived  
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence  
CC selected from those given in ABL70627 to ABL76833. The cdps sequences  
CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)  
CC can be used for determining altered gene expression, to recover  
CC regulatory elements and to follow inheritance of desirable  
CC characteristics through hybrid breeding programs. (I) are also useful  
CC in the evaluation, and alteration of desired characteristics associated  
CC with growth and development, disease resistance, environmental  
CC adaptability, quality and yield, and as molecular markers for studying  
CC inheritance of multigenic traits in a plant breeding program. (I) can be  
CC used to produce a tassel-specific profile of gene transcription, a  
CC transcript image, to clone regulatory elements for use in transformation  
CC vectors, to express a polypeptide, to identify, isolate or extend  
CC identical or related corn tassel nucleic acid sequences from DNA  
CC libraries, in nucleic acid hybridisation or amplification technologies,  
CC as query sequences to determine homology of known sequences, as probe  
CC for use in Southern or Northern hybridisation, and to identify the  
CC presence of and/or to determine the degree of similarity between two  
CC (or more) nucleic acid sequences.  
XX  
S0 Sequence 284 BP; 66 A; 55 C; 75 G; 86 T; 2 other;  
Query Match 5.9%; Score 96.8; DB 24; Length 284;  
Best Local Similarity 61.3%; Pred. No. 1.4e-16;  
Matches 155; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
OY 1177 TTATCTTAGAACCAAGCTTCTGGGCTGGTGGCTATGTCGATCACTGCTTCATGC 1236  
DB 1 TTTTGTTTGGGAGCCATTGACTCCTCTTATGACCAAGTTGGGATGACGCTTGGCTTG 60  
OY 1237 TAGCTGTGGTGGAGCTATACAGAAATATCTTTCGAAATCCACAATAACGCTCTCT 1296  
DB 61 CGGAGTCTATGTGGAGCAATGAGAACATTTTCAGTAAGAGCAATAAAGCTCTGT 120  
OY 1297 TTGACTCACTAAAGAAATGCGCTATATCCCTTTGACCAAGAGCAAAAGTCAAGTGA 1356  
DB 121 TTGATCTCTTGCNAAGAGATGCAATACATTTCTTGGATGAGAGATGAAGTGAAGTGA 180  
OY 1357 AGCGTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1416  
DB 181 AAGGCGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
OY 1417 AAGGTTGGCTCGT 1429  
DB 241 AGTTCATGATCTCT 253  
RESULT 10  
ABL76356  
ID ABL76356 standard; cDNA: 246 BP.  
XX  
AC ABL76356;  
XX

DT 14-MAY-2002 (first entry)  
XX  
DE Corn tassel-derived polynucleotide (cdps) SEQ ID NO: 5730.  
XX  
KW Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs;  
KW inheritance; characteristic; growth; development; disease resistance;  
KW environmental adaptability; quality; yield; molecular marker;  
KW multigenic trait; plant breeding; corn tassel; gene; ss.  
XX  
OS Zea mays.  
XX  
PN US2001051335-A1.  
XX  
PD 13-DEC-2001.  
XX  
PF 16-APR-1999; 9905-0294093.  
XX  
PR 21-APR-1998; 98US-082567P.  
XX  
PA (LALG/) LALGUDI R V.  
PA (ITOL/) ITO L Y.  
PA (SHER/) SHERMAN B K.  
XX  
PI Laljudi RV, Ito LY, Sherman BK;  
XX  
DR WPI; 2002-163647/21.  
XX  
PT Novel purified corn tassel-derived polynucleotide useful for  
PT determining altered gene expression, to recover regulatory elements and  
PT to follow inheritance of desirable characteristics through hybrid  
PT breeding programs  
XX  
PS Claim 1; SEQ ID 5730; 201pp; English.  
XX  
CC The present sequence describes a purified corn tassel-derived  
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence  
CC selected from those given in ABL70627 to ABL76833. The cdps sequences  
CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)  
CC can be used for determining altered gene expression, to recover  
CC regulatory elements and to follow inheritance of desirable  
CC characteristics through hybrid breeding programs. (I) are also useful  
CC in the evaluation, and alteration of desired characteristics associated  
CC with growth and development, disease resistance, environmental  
CC adaptability, quality and yield, and as molecular markers for studying  
CC inheritance of multigenic traits in a plant breeding program. (I) can be  
CC used to produce a tassel-specific profile of gene transcription, a  
CC transcript image, to clone regulatory elements for use in transformation  
CC vectors, to express a polypeptide, to identify, isolate or extend  
CC identical or related corn tassel nucleic acid sequences from DNA  
CC libraries, in nucleic acid hybridisation or amplification technologies,  
CC as query sequences to determine homology of known sequences, as probe  
CC for use in Southern or Northern hybridisation, and to identify the  
CC presence of and/or to determine the degree of similarity between two  
CC (or more) nucleic acid sequences.  
XX  
S0 Sequence 246 BP; 58 A; 49 C; 63 G; 76 T; 0 other;  
Query Match 5.2%; Score 85.6; DB 24; Length 246;  
Best Local Similarity 60.2%; Pred. No. 1.6e-13;  
Matches 142; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
OY 1093 TCATTGTAATTTGATGATTAACGTGAGCCCTAGTCACTCTGTCTGCTCTAA 1152  
DB 7 TCCACAGAAATTTGGGGGAGTGCAGTACATACGCCCTGCAATGTACTCCCTCA 66  
OY 1153 CAGGTATCGTTTCTTGGCTCTTGTATCTTTAGAACCAAGCTTGGGCTGGCTGCTGA 1212  
DB 67 CAGAGTTGGGCTTCTTCATCTATTTGTTGGTGCAGCCATTCCTCTTATGACCA 126  
OY 1213 TGTTCGGTAAACACCTCTCATGCTAGCTGTTGTCGGAGCTATACAGAAATTTCTT 1272  
DB 127 AGTTGGGATACACCTTGTCTGGGAGCTATGTTGGAGCAATGACAAATTTTCA 186



CC (II) and their complementary sequences, as primer oligonucleotides for  
CC the amplification of the 350 sequences, (II) and/or their complements and  
CC as oligomer probes for detecting the cytosine methylation state and/or  
CC single nucleotide polymorphisms (SNPs).  
CC Note: the sequence data for this patent did not form part of the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.

PR	20-OCT-2000	2000US-0241876
PR	20-OCT-2000	2000US-0241221
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250391
PR	01-DEC-2000	2000US-0251160
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251988
PR	05-DEC-2000	2000US-0256719
PR	06-DEC-2000	2000US-0251479
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251858
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251899
PR	08-DEC-2000	2000US-0251990
PR	11-DEC-2000	2000US-0254097
PR	05-JAN-2001	2001US-0259678
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI: 2001-541565/60.	
DR		
XX		
PT	Nucleic acids encoding 3324 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system	
PT	cancers and metastases -	
XX		
PS	Disclosure: SEQ ID NO 10757; 1701pp + Sequence Listing; English.	
XX		
CC	The invention relates to novel genes (ABA11004-ABA21534) and proteins	
CC	(ABR1678-ABR18001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful	
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast	
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	

[illegible]

Query	Subject	Score	Length	Mismatches	Gaps	Indels
629	TATTTCTTACTAGCTTCCTGCGTCGCACATTTGTTGGCGCTTAAGCTGAGGCTCCGT	688	10047	108	0	0
689	TTTCAGACGTGATGATCCCTTGGGGAATTTCTTACGCTTTTGAGGCGATGACTATGCT	748	10107	108	0	0
749	ATTCGACTTGTCTTATGCGCCAGTTACTGCTGCATCANTAGAACGATTTGACCGATCC	808	10167	108	0	0
809	TCCGCTTATTAATCCAGAAAGATCCAAAGGCGAATAA	847	10227	108	0	0
10227	TGTTTATATATTAATAAAGCATATATGAAGAAGTAA	10265				

RESULT 15  
 AAS61449  
 ID AAS61449 standard; DNA; 12669 BP.

29-JAN-2002 (first entry)

Human gene regulation-associated gene oligonucleotide #404.

Human; Gene regulation-associated gene; severe combined immunodeficiency; cardiac damage; inflammatory response; Haemophilia; Werner syndrome; asthma; HDR syndrome; congenital heart defect; Saelthre-Choitzen syndrome; renal disease; Preeclampsia; cardiac allograft vascular disease; colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour; immunostimulant; cardiac; antiinflammatory; coagulant; antislathmic; nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

Homo sapiens.

MO200177375-A2.

18-OCT-2001.

06-APR-2001; 2001WO-EP03968.

06-APR-2000; 2000DE-1019058.

07-APR-2000; 2000DE-1019173.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPig-) EPIDEMIOLOGY AG.

PI	Olek A, Plepenbrock C, Berlin K:
DR	WPI; 2002-017470/02.
PT	New nucleic acid sequences from chemically modified genes associated with gene regulation, useful for analysing cytosine methylations for diagnosis and therapy of diseases e.g. severe combined immunodeficiency disease -
PS	Disclosure; SEQ ID NO 412; 26pp; English.
XX	
CC	The invention relates to 224 nucleic acid sequences comprising at least 18 bases of a chemically pretreated gene associated with gene regulation selected from 43 known genes (or complementary sequences). The chemical pretreatment converts cytosine bases unmethylated at the 5-position to uracil or another base with hybridisation behaviour dissimilar to cytosine, to enable analysis of cytosine methylations. The DNA sequences, oligomers (or sets/arrays) and method are useful in the diagnosis of diseases (or predisposition to diseases) associated with gene regulation and in therapy of such diseases, by enabling analysis of the cytosine methylation patterns of such genes, kits are provided. They are especially useful in diagnosis and therapy of e.g. severe combined immunodeficiency disease, cardiac disorders, haemophilia, solid tumours and cancer, Werner syndrome, asthma, HDR syndrome, Saeltnre-Chotzen syndrome, renal disease, pre-eclampsia, graft versus-host disease. The present sequence is a sequence included in the sequence data for this specification and is associated with the human gene regulation-associated genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at <a href="http://wipo.int/pub/published_pct_sequences">ftp.wipo.int/pub/published_pct_sequences</a>
SQ	Sequence 12669 BP; 3610 A; 224 C; 2687 G; 6148 T; 0 other;
XX	
XX	Query Match 2.5%; Score 41.4; DB 24; Length 12669; Best Local Similarity 49.3%; Pred No. 1.4; Mismatches 108; Conservative 0; Indels 111; Gaps 0
DB	629 TATTTCTTACTAGACTTCGTCGTGCACAAATTGGCGCTCAAGATTGAGAGCTTCGT 688                                   10047 TTTTTTATTATGTTTGATGAAGCGTTGCTATACTTAATTAATAAAGTAGAATAAGTTAT 10106                                   QY 689 TTCTGAAGGTGATGATCCTTGGGGAATTTCTTACGCTTTTGATGGCTATGACTATGCT 748   DB 10107 TTTGTAGAAGGCGTGGATTTTAAATATTTATTTTATTTGCTTTTATTTTATTTTAGA 10166   QY 749 ATCTGACACTGTTCTCATGTGGCAGTATCAGTGATCAATGAAGACGATTTGACCAGTCC 808   DB 10167 ATTGAAGATTAGATGTTAATGCCGTTTAAAGAGATATGATATAGATTGATGACGGGTGA 10226   QY 809 TCGCTCTTATATCCAGANAATCCAAAGGGGAAAAA 847   DB 10227 TGTTTATATATTAATAAAGAGTAATAGAAAAGTAAGAAATA 10265 
RESULT 16	
ABK31523	
ID	ABK31523 standard; DNA; 12669 BP.
XX	
AC	ABK31523;
XX	
DT	23-APR-2002 (first entry)
DE	Signal transduction associated gene modified complementary DNA #183.
XX	
KM	Homo sapiens.
OS	Cpg island; signal transduction associated gene; cytosine methylation state; antitumour; cytostatic; mutant; ds.
XX	
SS	Synthetic.



PN WO200200926-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 29-JUN-2001; 2001WO-EP07472.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
PI WPI; 2002-147896/19.  
XX  
XX  
XX Oligonucleotide for diagnosis and therapy of diseases associated with  
PT signal transduction e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with signal transduction  
XX  
XX  
PS Claim 1; SEQ ID No 366; 24pp; English.  
XX  
XX The present invention relates to chemically modified DNA sequences of  
CC signal transduction associated genes. The DNA sequences are chemically  
CC modified using a solution of bisulphite, hydrogen sulphite or  
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers  
CC for detecting the cytosine methylation state (CpG islands) of these  
CC genes, and a method for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with signal transduction.  
CC The genomic DNA can be obtained from cells or cellular components which  
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,  
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from  
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver.  
CC histologic object slides, and all their possible combinations. The  
CC sequences of the invention are useful for the diagnosis and therapy of  
CC diseases associated with signal transduction e.g. solid tumours and  
CC cancer. AKK3158-ABK3155 represent chemically pretreated genomic DNA  
CC sequences of different genes associated with signal transduction, or  
CC their complementary sequences.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX  
SQ Sequence 12669 BP; 3610 A; 224 C; 2687 G; 6148 T; 0 other;  
Query Match 2.5%; Score 41.4; DB 24; Length 12669;  
Best Local Similarity 49.3%; Pred. No. 1.4; Indels 0; Gaps 0;  
Matches 108; Conservative 0; Mismatches 111;  
OY 629 TATTTCTTACTGCTTCTGTCGCAATGTTGGGCTCAAGTTGAGAGCTTCGT 688  
DB 10047 TTTTATATATTTGATGATAGGTTGATGATTTATTAATAAGATGATGATTTAT 10106  
OY 689 TTCTAAGAGTGTAGATCCTGGGGAATTTCTTACGCTTTTGGAGGCTATGATTTGT 748  
DB 10107 TTTTGAAGAGGCTGTGATTTTAAATTTTATTTTATTTTATTTTATTTTATTTT 10166  
OY 749 ATCTGAGACTGTTCTTATGAGCCAGTTACGTTGATCATTAAGATGATTAACGATCC 808  
DB 10167 ATTAAGATTTAGATTTGATTTGCTTTTACAGATATGATTAAGATTTTGAAGAGG 10226  
OY 809 TCGCTTCTATATCCAGAAAGATGCAAAAGGGAATAA 847  
DB 10227 TGTTTTATATATATAAAGATTAAGAAAGTAAGAAATA 10265  
RESULT 17  
AAK99539 standard; DNA; 1887 BP.  
XX  
XX AAK99539;  
XX  
XX 05-OCT-1999 (first entry)  
XX

DE Nucleic acid sequence from U. urealyticum.  
XX  
XX Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;  
KW human urogenital tract; pregnancy; neonatal disease; drug therapy;  
KW suppurative arthritis; ss.  
XX  
XX Ureaplasma urealyticum.  
OS WO9939007-A1.  
XX  
XX 05-AUG-1999.  
XX  
XX 29-JAN-1999; 99WO-US01972.  
XX  
XX 30-JAN-1998; 98US-0073189.  
XX  
XX (UABR-) UAB RES FOUND.  
PA  
PI Cassell GH, Chen EY, Glass JL, Glass JS, Heiner CR;  
PI Lefkowitz E;  
XX  
XX WPI; 1999-469343/39.  
XX  
XX  
XX Detection of Ureaplasma urealyticum using novel genes, probes and  
PT primers  
XX  
XX  
PS Claim 1; Page 31-32; 110pp; English.  
XX  
XX The present invention provides methods for the detection and diagnosis  
CC of Ureaplasma urealyticum infection. It provides novel genes  
CC (AAK99501-681) that can be used as a source of primers and probes for the  
CC detection and/or quantification of U. urealyticum in a biological  
CC sample. The probes that can be used in the method of the invention by  
CC forming target:probe complex is complementary to a region selected from  
CC one of the 181 nucleic acid sequences (AAK99501-681). U. urealyticum is  
CC an opportunistic pathogen of the human urogenital tract that is a  
CC significant cause of adverse pregnancy outcome, neonatal disease, and  
CC suppurative arthritis. As the infections are commonly asymptomatic, it is  
CC important to have specific and sensitive methods for detecting their  
CC presence in a patient. Also, as the pathogen has no current antibiotic  
CC directed specifically against it, it would be advantageous to isolate and  
CC detect gene sequences which are unique to it, and utilise these as a  
CC basis for diagnosis of U. urealyticum infection as well as to develop new  
CC and improved drug therapies. The present invention provides such novel  
CC polynucleotide sequences (AAK99501-681).  
XX  
XX  
SQ Sequence 1887 BP; 600 A; 188 C; 186 G; 913 T; 0 other;  
Query Match 2.5%; Score 41; DB 20; Length 1887;  
Best Local Similarity 51.2%; Pred. No. 0.78;  
Matches 149; Conservative 0; Mismatches 135; Indels 7; Gaps 2;  
OY 715 TTTCTTACGCTTTTGTGATGCTATGATGATGATGATGATGATGATGATGATGATGAT 774  
DB 824 TCTTTTGAATATTTTGTGTTTCAATACATAGATGATGATGATGATGATGATGATGAT 883  
OY 775 ACTGGGATCATTAAGAACGATATGACGATCGCTTC-----TATATCCAGAGAA 830  
DB 884 AATGATTTGAATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 943  
OY 831 ATGCAAAAGGAAAAAGAGCTTAACCTAAATGAATATGAAGATATGATGATGATGATGAT 890  
DB 944 TTTTAACTCTTGAATATATTTTGTAAATATTAATTAATTAATTAATTAATTAATTAAT 1003  
OY 891 CTTGATAGATGCTCTTATATTTCTTTTAACTGCTTGGTTATGCTTATGATGATGATGAT 947  
DB 1004 TCTTCAATATCTAATGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1063  
OY 948 TGCATTAATCTTAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 998  
DB 1064 TGTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1114



[illegible]

AAS45469/C  
ID AAS45469 standard; DNA; 8588 BP.  
AC AAS45469;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Chemically pretreated complementary DNA associated with cell cycle #87.  
XX  
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
RW human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;  
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds,  
KM PCR primer.  
XX  
OS Homo sapiens.  
XX  
PN W0200168911-A2.  
XX  
PD 20-SEP-2001.  
XX  
PE 15-MAR-2001; 2001WO-EP02945.  
XX  
PR 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K:  
XX  
DR WP1: 2001-602751/68.  
XX  
PT Designing primers and probes for analysing diseases associated with  
PT cytosine methylation state e.g., arthritis, cancer, aging,  
PT arteriosclerosis comprising fragments of chemically modified genes  
PT associated with cell cycle -  
XX  
XX  
Claim 1: SEQ ID No 174; 28pp; English.

XX  
PS Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
CC molecules associated with the cell cycle and specific PCR primers of the  
CC invention. The sequences are useful for detecting the methylation state  
CC of all CpG dinucleotides in a sequence and therefore for analysing  
CC associated diseases. By analysing cytosine methylations in the pretreated  
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
CC of existing diseases or the predisposition to specific diseases can be  
CC ascertained. The parameters may be compared to another set of genetic  
CC and/or epigenetic parameters, the differences serving as basis for  
CC diagnosis and/or prognosis events which are disadvantageous to patients.  
CC The sequences of the invention are useful for the diagnosis and therapy  
CC of HIV infection, neurodegenerative disorders, graft-versus host disease,  
CC aging, glomerular disease, Lewy body disease, arthritis,  
CC arteriosclerosis, solid tumours and cancers.  
XX  
SQ Sequence 8588 BP; 2142 A; 245 C; 2070 G; 4122 T; 9 other:

Query Match            2.4%; Score 39.2; DB 22; Length 8588;  
Best Local Similarity   66.7%; Pred. No. 4.8;  
Matches   56; Conservative   0; Mismatches   28; Indels   0; Gaps   0

OY     2   AAATTAATAAACCTCATAGAAATTAAACTATTTCAGACGGTAAATATGCACAANAAC   61  
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db     3945   AAAAAAAAAAATAAAAAAAAATAAATGAAGAATAACAAAATAAATAAATAAAAAAAA   3886  
                                |                      |                      |  
OY     62   CGAAGCAAAAACCCTTTGCCAAATT   85  
                                | | | | | | | | | | | | | | | | | | | | | |  
DB     3885   CGAAAAAAAACCGCGTAATAAATT   3862

XX	RESULT 22
ABR28325/C	
ID	ABR28325 standard; DNA: 8588 BP.
XX	
AC	
XX	ABR28325;
XX	
DT	23-APR-2002 (first entry)
XX	
DE	DNA transcription associated genomic DNA #100.
XX	
KW	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW	KW: cytosine methylation state; SNP: retroviral infection; gene; ds;
KW	single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW	viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW	immunological disorder; Werner syndrome; developmental disorder;
KW	psoriasis; Rieger's syndrome; neurologic disorder; erythropoiesis;
KW	neurodegenerative disorder; Maardenburg syndrome; Niemann-Pick disease;
KW	myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW	angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW	polyglutamine disorder; solid tumour.
XX	
OS	Unidentified.
XX	
PN	WO200192565-A2.
XX	
PD	06-DEC-2001.
XX	
PF	06-APR-2001; 2001WO-EP03973.
XX	
PR	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019173.
PR	30-JUN-2000; 2000DE-1032529.
XX	
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPiG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-090046/12.
XX	
PT	New nucleic acids or oligomers, useful for diagnosing or treating
PT	diseases associated with DNA transcription, e.g. immunological
PT	disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT	tumours or cancer -
XX	
PS	Claim 1; SEQ ID No 199; 32pp; English.
XX	
CC	The invention relates to a nucleic acid, which comprises a segment of the
CC	chemically pretreated DNA of genes associated with DNA transcription from
CC	one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC	or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC	to the chemically pretreated DNA of genes associated with DNA
CC	transcription. The set of oligomer probes are useful for detecting the
CC	cytosine methylation site and/or single nucleotide polymorphisms (SNPs)
CC	in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC	diagnosing or treating diseases associated with DNA transcription
CC	(particularly with the methylation status), e.g. adenosine deaminase
CC	deficiency, viral infection, retroviral infection, Sezary syndrome,
CC	haematological disorders, immunological disorders, Werner syndrome,
CC	tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC	neurological disorders, neurodegenerative disorders, Maardenburg
CC	syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC	infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC	disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC	or cancer. Sequences ABR28127-ABR28472 represent DNA transcription
CC	associated genomic DNA molecules of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
XX	Sequence 8588 BP; 2142 A; 245 C; 2070 G; 4122 T; 9 other:
XX	

Query Match 2.4%; Score 39.2; DB 24; Length 8588;

	Best Local Similarity	66.7%;	Pred. No.	4.8;	Mismatches	28;	Indels	0;	Gaps	0
	Matches	56;	Conservative	0;						
OY	2	AATAATMAAACCTCATCTGAGTACGAATAAATGATTTCAGAGCGGTAAATATGACAACAAAC	61							
Dd	3945	AAAAAAAAAAAAATTAATAAAAAAAAAAAAATTAACGCAAAAACCAGAAATATTAATTAATAAAAAAAAAA	3886							
OY	62	CGAGAGAAAACCTTTTGCAAAATTT	85							
Dd	3885	CGAAAAAAAAAACCCGGTAAAAAAATT	3862							
	RESULT 23									
ID	AAD22126	standard; DNA; 1231 BP.								
XX	AAD22126;									
AC	AAD22126;									
XX	12-FEB-2002	(first entry)								
DT	Lycopersicon esculentum sucrose transporter 1 (SUT1) gene 3' UTR.									
DE	Tomato; sucrose transporter 1; SUT1; transgenic plant; plant growth;									
XX	UTR; untranslated region; gene expression; regulatory element; ds.									
KM	Lycopersicon esculentum.									
XX	EP1149915-A1.									
PN	31-OCT-2001.									
PD	28-APR-2000; 2000EP-0109218.									
PF	28-APR-2000; 2000EP-0109218.									
XX	(FROM/) FROMMER W B.									
PA	Fischer W, Frommer WB, Hirner B, Lalonde S, Okumoto S, Tegeder M;									
PI	Ward J, Weise A;									
DR	WPI: 2002-012576/02.									
XX	Regulatory element for modifying gene expression in transgenic plants									
PT	e.g. enhance plant growth, comprises a promoter and a nucleotide									
PR	enhancer or repressor having a non-coding region of a sucrose									
PT	transporter gene -									
PS	Claim 9; Page 17; 35pp; English.									
XX	The invention relates to a regulatory element for the modification of									
CC	gene expression in transgenic plants. The regulatory element comprises a									
CC	promoter and a nucleotide sequence which is a nucleotide enhancer and/or									
CC	repressor sequence having a non-coding region of a sucrose transporter									
CC	(SUT) gene. The nucleotide enhancer sequence of is useful for enhancing									
CC	a promoter and nucleotide repressor sequence is useful for inhibiting a									
CC	expression of a plant gene which includes enhancing carbon dioxide uptake									
CC	in plant, enhance sucrose phloem loading capacity in plants, increase									
CC	sugar content or oil content in plant tissues or organs or increase									
CC	amino acid and/or protein content in plant tissues or organs. The									
CC	present sequence is Lycopersicon esculentum sucrose transporter 1									
CC	(SUT1) gene 3' untranslated region (UTR).									
XX	Sequence 1231 BP; 403 A; 156 C; 213 G; 459 T; 0 other:									
S0	Query Match	2.4%;	Score	39;	DB	24;	Length	1231;		
	Best Local Similarity	47.3%;	Pred. No.	2.3;						
	Matches	150;	Conservative	0;	Mismatches	165;	Indels	2;	Gaps	1.
OY	616	GATATGGAGCCTAATATTTCTTACTAGCTTCGTCGCCTGCCATTTGTTGGCTTCAAAGT	675							
Dd	184	GTTTCCTTTTAAATTTCTTCCAAACAAATGGGGTAGTGCTAGCTTTTTCAATTTGTGTG	243							

```

OY 676 TCAGACGCTCCGTTTCGAGGTGATGCTTTGGGGAATTTCTTACGCTTTTGATG 735
Db 244 TACCAAGAAAGATGTTCCAGTATTAACATGCTGGCAATGCTTTGACCTTTACTA 303
OY 736 CATACGATTTATCTGAGCTTTCTTATGCGCAGTACTGCGATCATTAAGACG 795
Db 304 TTTTATTTTATTTTGTAAATTTCTTCTTGGAAGTAGGAAAAAAGAAAGTGT 363
OY 796 TATTACGATTCGCTCTTATATCCAGAAATGCAAAAGGGAAGAAAGTGT 855
Db 364 TGGAGGTGAAGTGGGATCTTAATTGG--CTTGCTAAGAGGAGATTTATGAAGA 421
OY 856 AACCTAAATGAATGAAGATGCTCTCTATCTGATGATCTCTTATTTCTTT 915
Db 422 ACTTGTACTTAAAGTGAATCAAAATGATCTATCTATATCATCAATGTTCTAT 481
OY 916 TATTACGCTCTTGCT 932
Db 482 TTTTGTGTTCTTGCTT 498

```

## RESULT 24

AAV82021/C  
ID AAV82021 standard; DNA: 2718 BP.

AAV82021;

21-JUN-1999 (first entry)

XX Moraxella catarrhalis VH19 lbpB gene.

XX Lactoferrin receptor; lactoferrin binding protein; LBP1; LBP2;  
KW lbpA gene; lbpB gene; ORF3; infection; otitis media; sinusitis;  
KW conjunctivitis; pneumonia; bronchitis; tracheitis; emphysema;  
KW diagnosis; therapy; vaccine; Branhamella catarrhalis; ss.

XX Moraxella catarrhalis.

XX Key Location/Qualifiers  
FH 1, 2718  
FT CDS

```

FT //tag= a
FT /transl_except= (pos:844..846, aa:Thr)
FT /transl_except= (pos:847..849, aa:Thr)
FT /transl_except= (pos:850..852, aa:Thr)
FT /transl_except= (pos:853..855, aa:Thr)
FT /transl_except= (pos:856..858, aa:Thr)
FT /transl_except= (pos:859..861, aa:Thr)
FT /transl_except= (pos:862..864, aa:Thr)
FT /transl_except= (pos:868..870, aa:Thr)
FT /transl_except= (pos:871..873, aa:Thr)
FT /transl_except= (pos:874..876, aa:Thr)
FT /transl_except= (pos:877..879, aa:Thr)
FT /transl_except= (pos:880..882, aa:Thr)
FT /transl_except= (pos:886..888, aa:Thr)
FT /transl_except= (pos:891..891, aa:Thr)
FT /transl_except= (pos:892..894, aa:Thr)
FT /transl_except= (pos:895..897, aa:Thr)

```

XX W09855606-A2.

XX 10-DEC-1998.

XX 02-JUN-1998; 98WO-CA00544.

XX 08-MAY-1998; 98US-0074658.

XX 03-JUN-1997; 97US-0867941.

XX (CONN-) CONNAUGHT LAB LTD.

XX Du R, Klein MH, Loosmore SM, Wang Q, Yang Y;

XX WPI; 1999-070266/06.

XX P-PSDB; AAW89421.

XX Lactoferrin receptor genes from Moraxella, especially M. catarrhalis  
PT - useful to diagnose Moraxella infection e.g. to detect otitis media  
PT due to M. catarrhalis infection and to immunise against such  
PT infections

PS Claim 8; Fig 16A-O; 202pp; English.

XX This polynucleotide comprises the lactoferrin binding protein lbpB  
CC gene of Moraxella catarrhalis (Branhamella catarrhalis) strain VH19.  
CC It encodes lactoferrin binding protein 2 (lbp2, see AAW89421). The  
CC genes and DNA sequences of the lactoferrin receptor (lfr) locus of  
CC Moraxella are useful for diagnosis, immunisation, and the  
CC generation of diagnostic and immunological reagents. Immunogenic  
CC compositions, including vaccines, based upon expressed recombinant  
CC lbp1 and/or lbp2 and/or ORF3, portions of these, or their  
CC analogues, can be prepared for prevention of diseases caused by  
CC Moraxella. M. catarrhalis is a causative agent of otitis media and  
CC has been associated with sinusitis, conjunctivitis and inflammatory  
CC diseases of the lower respiratory tract, such as pneumonia, chronic  
CC bronchitis, tracheitis and emphysema.

XX Sequence 2718 BP; 958 A; 585 C; 547 G; 628 T; 0 other;

## Query Match

Best Local Similarity 57.4%; Score 38.8; DB 20; Length 2718;  
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```

OY 1079 TGTGTGTGTACGATTCGTAATTTGGATGTTAACTGAGCCCTAGTCACTCCTGT 1138
Db 1231 TATTGTGTTATCATCAATCAAAAGATTGGCTGCGAGTAAGGGTTCGCAATTAACCT 1172
OY 1139 CATGTTCTCCCAACAGGATGTTTCTTCCTCTGTTATCTTTAGAACCAAGCTTC 1198
Db 1171 TGGCGTCAATCTCATGACATCGGCTTCATTGACCTTTTGTGGTATCAAACTGACTTC 1112
OY 1199 TG 1200
Db 1111 TG 1110

```

## RESULT 25

AAH48024/C  
ID AAH48024 standard; DNA: 14041 BP.

AAH48024;

XX 18-SEP-2001 (first entry)

XX Internal control B19c #1.

XX Internal control; ss.

XX Parvovirus.

XX W0200146463-A2.

XX 28-JUN-2001.

XX 20-DEC-2000; 2000WO-EPI2996.

XX 22-DEC-1999; 99AT-0002170.

XX (BAXT ) BAXTER AG.

XX Zimmermann K, Turecek P, Schwarz H, Rieger M;

XX WPI; 2001-408658/43.

XX Internal standards useful for nucleic acid amplification assays,  
PT comprises a synthetic nucleic acid made by non-recombinant techniques

XX



PI Tectelin H;  
XX  
XX WPI: 2002-352536/38.  
DR P-PSDB: ABP28137.  
XX  
PT New Streptococcus protein for the treatment or prevention of infection  
PT or disease caused by Streptococcus bacteria, such as meningitis, and  
PT for detecting a compound that binds to the protein -  
XX  
PS Claim 7; Page 3706; 4525bp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
CC the specification. The proteins have antibacterial and anti-inflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.  
XX  
SQ Sequence 924 BP; 280 A; 159 C; 211 G; 274 T; 0 other;  
XX  
Query Match 2.4%; Score 38.6; DB 24; Length 924;  
Best Local Similarity 49.3%; Pred. No. 2.6;  
Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0;  
OY 558 GGATTGCTAATGAATACAAATCCAGAGCAAGCGTTCTACGCTCTTTTCGT 617  
DB 181 GAATTTGATGCTAAATCAAGAGATTGATGACAGAAATCGCTTCATCTTCACGT 240  
OY 618 ATCGAGCTAATATTTCTTACTAGCTTCTGCTGCAATTTGTTGGCTTCAAGTTG 677  
DB 241 CGTGAAGTTTGAAGAAATCAGCAGCTGCAGCAGTAAAGATTTCACAAACATTGAA 300  
OY 678 AGAGCTTCGCTTTCGAAGGTAGATCCTTGGGCAATTTCTTACGCTTTTGGATGCT 737  
DB 301 GTAGGTTCAAGTTGTAAGTGTGCTGTTTACAGACTTTGGTCTTTCATTGAC 360  
OY 738 ATGACTATTGTATCTGAGCTGTTTC 762  
DB 361 CTTGGTGGTGTGATGAGACTTGTAC 385  
RESULT 28  
AAH52921  
ID AAH52921 standard; DNA: 1176 BP.  
XX  
AC AAH52921;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1235.  
XX  
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KW vaccination; endocarditis; ds.  
XX  
OS Staphylococcus epidermidis.  
XX  
XX WO200134809-A2.  
XX  
XX 17-MAY-2001.  
XX  
XX 09-NOV-2000; 2000WO-US30782.  
XX

PR 09-NOV-1999; 99US-0164258.  
XX  
XX (GLAXO) GLAXO GROUP LTD.  
XX  
XX  
PI Kimerly WJ;  
XX  
XX WPI: 2001-316495/33.  
DR P-PSDB: AAG82071.  
XX  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 8; Page 354; 2188bp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG8154 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 1176 BP; 342 A; 164 C; 219 G; 451 T; 0 other;  
XX  
Query Match 2.4%; Score 38.6; DB 22; Length 1176;  
Best Local Similarity 49.3%; Pred. No. 2.9;  
Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0;  
OY 1054 TAGTATCCCTACTATCATCTATTTGTTGGGTAGCTCATCTGTAATTTGGATGCT 1113  
DB 827 TACTATTGTGTAACATCATCATATTGTTAGCTGTTATTCAAATTCATATTAGCTTAA 886  
OY 1114 TAACGTGAGCCCTGACTCCTGTCATGCTCTCCTAACAGATATGTTTCTTCGCTC 1173  
DB 887 TCATTTGCTTCTATTCGGTGTGATGATTTGGTACACACCGTACTTAATAGCAAA 946  
OY 1174 TTGTTATCTTTAGAAACCAAGCTTCTGGGCTGCTGCTATGTTGGTACACTCCCTCA 1233  
DB 947 TTATCTTAAGTGCMAAAGAACACACTTCTTGCAAGTAGCTTAGCTGCTTATTTTCA 1006  
OY 1234 TCGTAGCTGTGTTGTCGAGCTAT 1258  
DB 1007 ATGTTGCTAATTTTCTTGTGTCAT 1031  
RESULT 29  
ABN70725  
ID ABN70725 standard; DNA: 1398 BP.  
XX  
AC ABN70725;  
XX  
DT 01-JUL-2002 (first entry)  
XX  
DE Streptococcus polynucleotide SEQ ID NO 9363.  
XX  
XX Streptococcus polynucleotide  
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
KW anti-inflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX  
OS Streptococcus agalactiae.  
XX







QY 727 TTTTGATGCTATGATCTATGCTGACCTTGTTCTTATGCGCAGTACTGATGATCA 786  
11 111 11 111 111 11 111 111 111 111  
Db 1548 TTGAGAGCGAATTATGTTAGTGTGATTAATTTATTTTCGATTTAGGATGAGATG 1607  
QY 787 ATACAGACGATTT 799  
111111 111111  
Db 1608 TGAAGAAAGGAAT 1620

RESULT 33  
ABU92259  
ID ABU92259 standard; DNA; 7695 BP.  
XX  
AC ABL92259;  
XX  
DT 01-JUL-2002 (first entry)  
XX

De Chemically treated DNA repair gene fragment complementary to#34.  
XX  
XX DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3;  
KM PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;  
KM DDTL1; FANCB; XRC8; ataxia telangiectasia; aging; Bloom's syndrome;  
KM Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;  
KM immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;  
cancer; ds.  
XX  
XX Unidentified.  
OS  
XX  
PN W0200181622-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 06-APR-2001; 2001WO-EP03972.  
XX

PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX

PA (EPIC-) EPIDENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-034446/04.  
XX

PT New nucleic acid derived from genes associated with DNA repair, useful  
PT for diagnosis, e.g. of ataxia telangiectasia, by determination of  
PT cytosine methylation -  
XX

PS Claim 1; SEQ ID NO 68; 25pp + sequence listing; English.  
XX

CC The invention relates to nucleic acids containing a sequence of at least  
CC 18 nucleotides of chemically treated DNA of genes associated with DNA  
CC repair, and their complements. The invention also relates to nucleic  
CC acids comprising at least 18 base pairs of the chemically pretreated DNA  
CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,  
CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, MGMT, MSH2, NUDT1, TDG, INPPL1,  
CC RFC4, DDTL1, FANCB, or XRC8. Nucleic acids of the invention and related  
CC oligomers, are useful for diagnosis of diseases associated with gene  
CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,  
CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,  
CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours  
CC and cancer, particularly by determining status of cytosine methylation  
CC and/or by detecting single-nucleotide polymorphisms. Determination of  
CC individual methylation patterns may allow development of individualised  
CC therapies. The sequences given in records ABU92192-ABU92335 represent  
CC chemically pre-treated DNA fragments from genes associated with DNA  
CC repair, and their complements.  
CC Note: The sequence data for this patent is not represented in the  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.  
XX  
SQ Sequence 7695 BP; 2111 A; 268 C; 1887 G; 3429 T; 0 other;

Query Match 2.4%; Score 38.6; DB 24; Length 7695;  
Best Local Similarity 43.3%; Pred. No. 6.7;  
Matches 179; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY 560 ATTGCTAATGCAATTCACAAAATCCACGAAAGCAAGCGTTCTACGCTCTTTGGCTAT 619  
1111 11111111 1111 1111 1111 1111 1111 1111 1111 1111  
Db 5182 ATTTTAAATTAATTAATTTAGAAAAGAAAAGAAAATATTTATTTATTTTGT 5241

QY 620 CGAGAGCTAATATTTCTTACTAGCTTCGCGTCGCAATGTTGGGCTCCAAAGTGAG 679  
1111 11111111 1111 1111 1111 1111 1111 1111 1111 1111  
Db 5242 TGAATAGCATATATTTATTAATTTTGTGTTTTAGTAGTGTTGTTTTAAATGTTAT 5301

QY 680 AGCTTCGCTTCTGAGAGCTAGATCCTTGGGCAATTTCTTACGCTTTGATGCTAT 739  
1111 11111111 1111 1111 1111 1111 1111 1111 1111 1111  
Db 5302 ATGTTTAAATATAGTTTTTTTTTTTAAAGAGATTTTTTAAAGCTTTTATGATTTTTT 5361

QY 740 GACTTATTTGATCTGACCTGTTCTTATAGCCAGTTACTGTGATCAATAGAAGCTATT 799  
1111 11111111 1111 1111 1111 1111 1111 1111 1111 1111  
Db 5362 GGTAAAGAAAGAAAGAAATGTTGATATATATTTGAAATTTTATATATGTAACGATT 5421

QY 800 GACGATCCCTCGCTTATATATCCAGAAAGATCCAAAGGGAAAGAGTGCTAACC 859  
1111 11111111 1111 1111 1111 1111 1111 1111 1111 1111  
Db 5422 TGGGAGGTAATTAATTAATTTTAAATTTAAAGTTACGTTTTTAAATAGTAAATTAG 5481

QY 860 TAAATGAATATGAAGATAGCTCTCTATCTGATAGATCCTTATATCTTTATT 919  
11111111 1111 1111 1111 1111 1111 1111 1111 1111 1111  
Db 5482 TAAATATATATTTGTTATTTTATTTTATTTTATGATTAATGTTATGGAGATATTTAA 5541

QY 920 AACTCTCTTGTTGTTATTCCTATAGTATTTGATTAACCTTAATGCAAGTGACTT 972  
1111 11111111 1111 1111 1111 1111 1111 1111 1111 1111  
Db 5542 ATGTGAATTTGCAATATATTAATTAATTTGATTAATAAATTTGATAGATTAATT 5594

RESULT 34  
ABL33537  
ID ABL33537 standard; DNA; 7695 BP.  
XX  
XX ABL33537;  
XX  
DT 26-MAR-2002 (first entry)  
XX

De Human immune system associated gene SEQ ID NO: 1510.  
XX

KM Human; immune system disease; cytosine methylation; anasthmatic;  
KM arteriosclerotic; antianemic; cytostatic; neotropic;  
KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KM antineumatic; antiarthritic; antidiabetic; antipsoriatic;  
KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KM acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KM gene; ds.  
XX

OS Homo sapiens.  
XX  
PN W0200200928-A2.  
XX  
PD 03-JAN-2002.  
XX

PF 02-JUL-2001; 2001WO-EP07537.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX

PA (EPIC-) EPIDENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130909/17.  
XX

PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -



XX	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
PI	Telford J., Massignani V., Margarit Ros YI, Grandi G, Fraser C;
P1	Tetelin H;
XX	
DR	WPI; 2002-352536/38.
XX	
PT	New Streptococcus protein for the treatment or prevention of infection
or	disease caused by Streptococcus bacteria, such as meningitis, and
PT	for detecting a compound that binds to the protein -
PS	Claim 8; Page 4196-4488; 4525pp; English.
XX	
CC	The invention relates to a protein (ABP25413-ABP30895) from group B
CC	streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC	the specification. The proteins have antibacterial and antiinflammatory
CC	activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
CC	antibodies that bind (1) are used in the manufacture of medicaments for
CC	the treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (1) are used to detect Streptococcus in a
CC	biological sample. (1) is used to determine whether a compound binds to
CC	(1). A composition comprising (1) or a nucleic acid encoding (1), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (1) may be used to recombinantly produce (1) and may be
CC	used in gene therapy. Antibodies to (1) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
CC	Streptococcus proteins.
XX	
SQ	Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;
	Query Match            2.4%: Score 38.6; DB 24; Length 2155561;
	Best Local Similarity    49.3%: Pred. No. 79;
	Matches    101; Conservative    0; Mismatches    104; Indels         0; Gaps         0;
OY	558 GGATTTGCTAATGAATTACAAAAATCCAGCAGAAGCGTTTCTACGCTCTTTTCGGT 617
Db	1124096 GAATTGTGATCTCTAAATCAAGGAAGCTTGATGCACGAAATAATCGCTTATCTTCACGT 1124037
OY	618 ATCGAGCTATATTCTTCTTACTAGCTTCGTGCTGCGAATTGTTGGGCTTCAAAGTTG 677
Db	1124036 CGGGAATCTTGTGAAGATCAGCAGCGCACGTAAGAAGATTTTCTCAAAACATTGAA 1123977
OY	678 AGAGCTTCCGTTTCTGGAAGCTGTAGATCCCTGGGGAATTTCTTACGCTTTTGTAGGCT 737
Db	1123976 GTAGGTTCACTTGTGAACCTGGTAAGTTGCTGCTTTAACGAAGCTTTGGTGCCTTTCATTGAC 1123917
OY	738 ATGACTATTGTATCTGCAGCTTGTC 762
Db	1123916 CTGCGTGTGCTTGTATGATGACCTTGAC 1123892
RESULT 37	
ID	ABN91748
ABN91748	standard; DNA; 1269 BP.
XX	
AC	ABN91748;
XX	
DT	24-JUL-2002 (first entry)
XX	
DE	Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO.1211.
XX	
KM	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW	antibacterial; gene therapy; gene; ds.
XX	
OS	Staphylococcus epidermidis.
XX	
NN	US6380370-B1.

[illegible]



PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 XX  
 PS Claim 1; SEQ ID NO 1642; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SQ Sequence 5347 BP; 1429 A; 60 C; 1153 G; 2705 T; 0 other;

Query Match 2.3%; Score 38.2; DB 24; Length 5347;  
 Best Local Similarity 52.8%; Pred. No. 7.3;  
 Matches 105; Conservative 0; Mismatches 93; Indels 1; Gaps 1;  
 QY 737 TATGACTATGTATCGACTGTCTTATGCGCCAGTTACTGCGATCAATPAGACGT 796  
 Db 1653 TATGGAATGATATATAGCTTTTATGGAATTTAATTATAGATTTTGTGTGG 1712  
 QY 797 ATTGACCGCTCCTCTATATCCAGAGAAATGCAAAAGCGAAAGTGCTAA 856  
 Db 1713 TTTTGAAGATTTTGTGTTTATAGATTAAGTAATTAATTAATGATGATGTA 1772  
 QY 857 ACCTAAATGAATAGAAGATAGCTTCTATCTTGATAGATCTCCTATATATCTTT 916  
 Db 1773 TGAGAAATTAGTATGAATATAATTTTTTTTGTGTTTATGATATGCGAA-TTTT 1831  
 QY 917 ATTAACCTCTTGTTATT 935  
 Db 1832 ATTAATTTTTTTATTATT 1850

Search completed: January 21, 2003, 19:15:44  
 Job time : 4255 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 19:33:03 : Search time 253 Seconds  
(without alignments)  
14571.230 Million cell updates/sec

Title: US-09-869-433-1  
Perfect score: 1637  
Sequence: 1 gaataaanaactacagaa.....gttgatccctactctgtt 1637

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues  
Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : N.Geneseq\_101002:\*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	1637	100.0	1637 21 AAA30922 C. pneumoniae ATP/
2	1535	93.8	1230025 20 AAX91990 Nucleotide sequenc
3	1484	90.7	273254 21 AAC81914 Chlamydia pneumoni
4	39	2.4	1038602 20 AAZ01425 Complete genome se
5	23	1.4	43 21 AAA30923 PCR primer for C.
6	22	1.3	786 22 AAL23777 Human breast cance
7	21	1.3	2139 22 AAS30049 Human lung antigen
8	21	1.3	4673 22 AAS30048 Human lung antigen
9	21	1.3	5032 24 AAS94838 Human DNA sequence

10	21	1.3	8220	16	AAO83529	P. falciparum Proj
11	21	1.3	8220	18	AAT72897	Plasmodium Proj3 g
12	21	1.3	8220	21	AAZ98286	P. falciparum Proj
13	21	1.3	19124	18	AAT72882	Plasmodium var-7 g
14	21	1.3	19124	21	AAZ98287	Plasmodium var-7 p
15	20	1.2	20	20	AAX95295	PCR primer used to
16	20	1.2	20	20	AAX95297	PCR primer used to
17	20	1.2	20	20	AAX95298	PCR primer used to
18	20	1.2	20	20	AAX92828	PCR primer used to
19	20	1.2	20	20	AAX92823	PCR primer used to
20	20	1.2	20	20	AAX92826	PCR primer used to
21	20	1.2	1770	22	AAF61373	A. thaliana chloro
22	20	1.2	2287	22	AAK77984	Nucleotide sequenc
23	20	1.2	33780	22	AAH24652	Human cDNA clone (
24	20	1.2	136284	24	ABK83575	Polynucleotide seq
25	20	1.2	562	22	AAH18882	Human neuroblastom
26	19	1.2	644	20	AAX20940	DNA encoding novel
27	19	1.2	764	22	AAI96504	Haemophilus influe
28	19	1.2	790	23	AA574111	DNA sequence upstr
29	19	1.2	1001	22	AAF91436	Human prostate exp
30	19	1.2	1001	24	ABK37818	Human microglobuli
31	19	1.2	1362	23	ABV25656	Human glycosyl-pho
32	19	1.2	1604	24	ABA02319	Human cDNA sequenc
33	19	1.2	1689	24	ABL41351	Drosophila melanog
34	19	1.2	1833	22	AAH15049	Drosophila melanog
35	19	1.2	2611	23	ABL12876	Drosophila melanog
36	19	1.2	3052	23	ABL04386	DNA encoding SCAN
37	19	1.2	3752	24	ABK13001	Drosophila melanog
38	19	1.2	4507	23	ABL24540	Drosophila melanog
39	19	1.2	5586	23	ABL11410	Drosophila melanog
40	19	1.2	8034	23	ABL03008	Drosophila melanog
41	18	1.1	384	22	AAS23970	Human ovarian PC
42	18	1.1	384	24	ABK39720	cDNA encoding lung
43	18	1.1	396	22	AAS24335	Human ovarian PC
44	18	1.1	404	22	AAI01875	Human reproductive
45	18	1.1	404	23	ABL97168	Human testicular a
46	18	1.1	426	14	AAO60207	Human brain Expres
47	18	1.1	435	22	AAK30727	Human bone marrow
48	18	1.1	459	24	ABK97583	Human prostate spe
49	18	1.1	481	22	ABA54121	Human fetal liver
50	18	1.1	481	22	ABA23868	Probe #2334 for ge
51	18	1.1	481	22	AAK27835	Human bone marrow
52	18	1.1	481	22	AAI02330	Human testicular a
53	18	1.1	484	22	AAH82924	Human ovarian tumo
54	18	1.1	486	22	AAH82533	Human ovarian tumo
55	18	1.1	687	21	AAA89558	Exo3-8 nucleotide
56	18	1.1	994	22	AAI06036	Human reproductive
57	18	1.1	994	23	ABL96601	Human testicular a
58	18	1.1	994	23	ABL96037	Human reproductive
59	18	1.1	995	23	ABL96602	Human testicular a
60	18	1.1	1298	22	AAH29860	C albicans apoptos
61	18	1.1	1681	22	AAF58306	Human GTP-binding
62	18	1.1	1876	22	AAF45074	Human secreted pro
63	18	1.1	1876	22	AAF33041	Human secreted pro
64	18	1.1	1876	24	ABO54209	Human ovarian anti
65	18	1.1	2001	22	AAK52772	Human polynucleoti
66	18	1.1	2043	22	AAK51788	Human polynucleoti
67	18	1.1	2677	22	AAI93897	Human stomach can
68	18	1.1	2677	22	AAH17908	Human stomach can
69	18	1.1	2727	23	ABV24810	Human prostate exp
70	18	1.1	2727	23	ABV25689	Human prostate exp
71	18	1.1	2748	22	AAI93884	Human stomach can
72	18	1.1	2748	22	AAH17695	Human cDNA sequenc
73	18	1.1	3022	20	AAZ20677	P. vulgaris chondr
74	18	1.1	3343	23	ABL20092	Drosophila melanog
75	18	1.1	3486	15	AAO71820	Chondroitinase ABC
76	18	1.1	3515	23	ABL05932	Drosophila melanog
77	18	1.1	3694	24	ABK83604	Human cDNA differe
78	18	1.1	3694	24	ABN95118	Gene #1616 used to
79	18	1.1	3699	24	ABK34876	Human cDNA encodin
80	18	1.1	3980	15	AAO74988	P. vulgaris chondr
81	18	1.1	3980	15	AAO74989	P. vulgaris chondr
82	18	1.1	3980	18	AAT58980	Proteus vulgaris c

83	18	1.1	3980	18	AAT58981	Mutated Proteus vu	
84	18	1.1	5327	22	AAS45357	Chemically pretrea	
85	18	1.1	5327	24	ABK28190	DNA transcription	
c	86	18	1.1	5591	23	ABL15092	Drosophila melanog
c	87	18	1.1	5591	23	ABL21058	Drosophila melanog
88	18	1.1	6065	22	ABL3262	Human musculoskele	
89	18	1.1	6240	23	ABK69060	Human BRP genomic	
90	18	1.1	6261	23	ABL14623	Drosophila melanog	
91	18	1.1	6519	15	AAQ74990	P. vulgaris chondr	
92	18	1.1	6519	18	AAT58982	Proteus vulgaris c	
93	18	1.1	10200	23	ABL14622	Drosophila melanog	
94	18	1.1	12768	23	ABL12988	Drosophila melanog	
95	18	1.1	14736	19	AAV52304	Streptococcus pneu	
c	96	18	1.1	16137	22	AAK87691	Human immune/haema
c	97	18	1.1	20445	24	AAS19906	Reference sequence
98	18	1.1	28420	22	AAK78286	Human Immune/haema	
c	99	18	1.1	34488	22	AAF97854	Human neuroblastom
100	18	1.1	40862	24	ABL34073	Human immune syste	

## ALIGNMENTS

```
RESULT 1
AAA30922 standard; DNA; 1637 BP.
ID AAA30922
AC AAA30922;
XX
XX 22-SEP-2000 (first entry)
DE C. pneumoniae ATP/ADP translocase coding sequence.
XX
XX ATP/ADP translocase; Chlamydia infection; diagnosis; therapy; ds.
XX
XX Chlamydia pneumoniae.
OS
XX
XX Key Location/Qualifiers
XX CDS 51..1598
XX /*tag= a
XX /*product= ATP/ADP_translocase
XX
XX MO200039157-A1.
XX
XX 06-JUL-2000.
XX
XX 22-DEC-1999; 99WO-CA01224.
XX
XX 28-DEC-1998; 98US-0114060.
XX 12-MAR-1999; 99US-0123967.
XX 30-JUN-1999; 99US-0141271.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J, Dunn P;
XX WPI: 2000-452368/39.
XX
XX P-PSDB; AAY90265.
XX
XX Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,
XX prevention and treatment of Chlamydia infection in mammals -
XX
XX Claim 2; Fig 1; 81pp; English.
XX
XX This sequence encodes the Chlamydia pneumoniae ATP/ADP translocase
XX of the invention. The protein, DNA encoding it, or a vaccine containing
XX the DNA or protein, are useful for diagnosing, preventing or treating
XX Chlamydia infection. The sequences can also be used in a method for
XX the detection of Chlamydia infection. Primers or probes derived from the
XX DNA sequence are useful in diagnostic tests for detecting Chlamydia
XX infection.
XX
XX Sequence 1637 BP; 431 A; 327 C; 325 G; 554 T; 0 other;
XX
XX
```

Query Match 100.0%; Score 1637; DB 21; Length 1637;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1637; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAAATAAAAA	ACTATCAGATATGA	AAATAAAGTATTT	CACAGGGTAA	ATATGACAAAA	60
DB	1	GAAATAAAAA	ACTATCAGATATGA	AAATAAAGTATTT	CACAGGGTAA	ATATGACAAAA	60
QY	61	CCGAGAAAAA	CTTTGGAAATATG	CGCTTTCTTG	GGCCGATACAT	ATCAGAGC	120
DB	61	CCGAGAAAAA	CTTTGGAAATATG	CGCTTTCTTG	GGCCGATACAT	ATCAGAGC	120
QY	121	TAAAGAAATT	CGCAATGTTCT	TAATGTTCTTG	ATTAATTAAT	TAATAGAGTGT	180
DB	121	TAAAGAAATT	CGCAATGTTCT	TAATGTTCTTG	ATTAATTAAT	TAATAGAGTGT	180
QY	121	TAAAGAAATT	CGCAATGTTCT	TAATGTTCTTG	ATTAATTAAT	TAATAGAGTGT	180
DB	121	TAAAGAAATT	CGCAATGTTCT	TAATGTTCTTG	ATTAATTAAT	TAATAGAGTGT	180
QY	181	TACGGATACA	AAAGACACTT	TATGTTGGAG	CTCTGTTCTG	TCGACAGGCAATAC	240
DB	181	TACGGATACA	AAAGACACTT	TATGTTGGAG	CTCTGTTCTG	TCGACAGGCAATAC	240
QY	241	CTTTCATCA	AGTTTGGCTGT	GTGCCCCGTG	CTATATCTT	ATGCTTATTTGCA	300
DB	241	CTTTCATCA	AGTTTGGCTGT	GTGCCCCGTG	CTATATCTT	ATGCTTATTTGCA	300
QY	301	AGCTAAGTA	ATATTTAAGTA	TAGACAGGCT	TATTTATG	ACATGGAACGCC	360
DB	301	AGCTAAGTA	ATATTTAAGTA	TAGACAGGCT	TATTTATG	ACATGGAACGCC	360
QY	361	TTTTCTTGG	CCCTGTCCCG	AGCTAATTTAT	CCGCTACG	CGATTTTACATCT	420
DB	361	TTTTCTTGG	CCCTGTCCCG	AGCTAATTTAT	CCGCTACG	CGATTTTACATCT	420
QY	421	AATTTGCT	AGCCGTTTAC	AGGCATCTAC	CTCCAGAT	TGCTAGACCTCG	480
DB	421	AATTTGCT	AGCCGTTTAC	AGGCATCTAC	CTCCAGAT	TGCTAGACCTCG	480
QY	481	TAAGAAC	TGACATTTCT	CGATTTTAT	GTACTGTG	AACATATGGGAA	540
DB	481	TAAGAAC	TGACATTTCT	CGATTTTAT	GTACTGTG	AACATATGGGAA	540
QY	541	TATCTCTA	TATGTTGGG	AGATTTGCT	ATATGAAAT	TACAAAATCC	600
DB	541	TATCTCTA	TATGTTGGG	AGATTTGCT	ATATGAAAT	TACAAAATCC	600
QY	601	TCTACGCT	TTTGGTATG	CGAGCTAAT	TATCTTAC	TACTCTGTCG	660
DB	601	TCTACGCT	TTTGGTATG	CGAGCTAAT	TATCTTAC	TACTCTGTCG	660
QY	661	TTTGGCTT	CAAGTTGAG	AGCTTCCG	TTTCTGAAG	GTGTAGATCCT	720
DB	661	TTTGGCTT	CAAGTTGAG	AGCTTCCG	TTTCTGAAG	GTGTAGATCCT	720
QY	721	TACGCTTT	TGATGCTAT	GACTATTTAT	CTGTGACT	TTTGTGCGCA	780
DB	721	TACGCTTT	TGATGCTAT	GACTATTTAT	CTGTGACT	TTTGTGCGCA	780
QY	781	GGATCAAT	TAAGAGCT	ATTTAGAC	CGATCCG	CTTATATATC	840
DB	781	GGATCAAT	TAAGAGCT	ATTTAGAC	CGATCCG	CTTATATATC	840
QY	841	GGAAAAAG	GTGCTTAA	ACCTTAAAT	GAATATGA	ATAGCTTCC	900
DB	841	GGAAAAAG	GTGCTTAA	ACCTTAAAT	GAATATGA	ATAGCTTCC	900
QY	901	CTCTTAT	ATTTCTTT	TATTAAC	TCTCTTG	GTATTTG	960
DB	901	CTCTTAT	ATTTCTTT	TATTAAC	TCTCTTG	GTATTTG	960
QY	961	TGCAAGT	ACTTGGAA	AGCTAG	CAACGCA	TATCTCAAT	1020
DB	961	TGCAAGT	ACTTGGAA	AGCTAG	CAACGCA	TATCTCAAT	1020



OY 1021 ACTTCATGGGAACTTCCTCTTGAGCTGCGCTAGTATCCGTAATTATCATGTATTTG 1080  
 DB 1021 ACTTCATGGGAACTTCCTCTTGAGCTGCGCTAGTATCCGTAATTATCATGTATTTG 1080  
 OY 1081 TTGGTGTACGCTATCTGTAATTTGGATGGTTAACTGAGCGCTAGTCACTCCTGTCA 1140  
 DB 1081 TTGGTGTACGCTATCTGTAATTTGGATGGTTAACTGAGCGCTAGTCACTCCTGTCA 1140  
 OY 1141 TGGTCTCTTACAGTATGCTTTTCTGCTCTTCTTATCTTAAACCAAGCTTCG 1200  
 DB 1141 TGGTCTCTTACAGTATGCTTTTCTGCTCTTCTTATCTTAAACCAAGCTTCG 1200  
 OY 1201 GGCTGTGCTATGTTGCTGTAACACTCTCTCATGCTAGCTGGTGGAGCTATAC 1260  
 DB 1201 GGCTGTGCTATGTTGCTGTAACACTCTCTCATGCTAGCTGGTGGAGCTATAC 1260  
 OY 1261 AGAATATCTTTCGAAATCCACAAATACGCTCTCTTACTCACTAAAGAAAGGCT 1320  
 DB 1261 AGAATATCTTTCGAAATCCACAAATACGCTCTCTTACTCACTAAAGAAAGGCT 1320  
 OY 1321 ATATCCCTCTTGACCAAGCAAAAGTCAAGGCTGCTATGATGATGTTGCGG 1380  
 DB 1321 ATATCCCTCTTGACCAAGCAAAAGTCAAGGCTGCTATGATGATGTTGCGG 1380  
 OY 1381 CCCGCTTCGAAATCAGAGAGAGCTTTATCCACACAGGTTTGCCTTATCTGTGAA 1440  
 DB 1381 CCCGCTTCGAAATCAGAGAGAGCTTTATCCACACAGGTTTGCCTTATCTGTGAA 1440  
 OY 1441 GATATGAGCTATGACCCCTTATCTTGCAAGTATCTTCTTTCATCATTTGATTTGGT 1500  
 DB 1441 GATATGAGCTATGACCCCTTATCTTGCAAGTATCTTCTTTCATCATTTGATTTGGT 1500  
 OY 1501 TGGTCTCTGACATAGTAAACAACATCTTCTTACGAGCTGCTTAAAGAACAG 1560  
 DB 1501 TGGTCTCTGACATAGTAAACAACATCTTCTTACGAGCTGCTTAAAGAACAG 1560  
 OY 1561 AAGTGTCTAAGAGATGACGCTGCTTCTCATAGAGTTGCTCTTACTCTTGT 1620  
 DB 1561 AAGTGTCTAAGAGATGACGCTGCTTCTCATAGAGTTGCTCTTACTCTTGT 1620  
 OY 1621 GATCCCTACCTGCTTTT 1637  
 DB 1621 GATCCCTACCTGCTTTT 1637  
 RESULT 2  
 AAX91990  
 ID AAX91990 standard; DNA: 1230025 BP.  
 AC AAX91990;  
 XX  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 XX  
 KM Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;  
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KM vaccine; neutralising epitope; ss.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO927105-A2.  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97RR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Grifffais R;

XX  
 DR WPI: 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 XX  
 PS Claim 1: Page 291-611; 1912pp; English.  
 XX  
 CC The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.  
 XX  
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;  
 Query Match 93.8%; Score 1535; DB 20; Length 1230025;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1635; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 GAATATAAAACATATCAGATAAGATAAAGTATTTCAGAGGTAAATATGACAAAA 60  
 DB 404320 GAATATAAAACATATCAGATAAGATAAAGTATTTCAGAGGTAAATATGACAAAA 404379  
 OY 61 CCGAAGAAAAACCTTTGGAAAAATTTGGCTCTTCTTGGCCGATACATCTACAGAC 120  
 DB 404380 CCGAAGAAAAACCTTTGGAAAAATTTGGCTCTTCTTGGCCGATACATCTACAGAC 404439  
 OY 121 TAAAGAACTTCCCAATGTTCTTATGTTCTCTGATATTAATTACTATACGGTGT 180  
 DB 404440 TAAAGAACTTCCCAATGTTCTTATGTTCTCTGATATTAATTACTATACGGTGT 404499  
 OY 181 TACGCATCAAAAGACACCTTATGTTGGAGACTCCTGATCTGTGCAAGAGCAATAC 240  
 DB 404500 TACGCATCAAAAGACACCTTATGTTGGAGACTCCTGATCTGTGCAAGAGCAATAC 404559  
 OY 241 CTTTCATCAAGTTTGGCTGTTGTCCTGTCCTATTAATCTTATGCTTATATGCAA 300  
 DB 404560 CTTTCATCAAGTTTGGCTGTTGTCCTGTCCTATTAATCTTATGCTTATATGCAA 404619  
 OY 301 AGCTAAGTAAATATTTAAGTAGACAGCCCTATTTATGCAAGTGGGACGCCCTTTTAA 360  
 DB 404620 AGCTAAGTAAATATTTAAGTAGACAGCCCTATTTATGCAAGTGGGACGCCCTTTTAA 404679  
 OY 361 TTTTCTTGGCCCTGTTCCGAGCTGAATTATCCGCTACGCGATGTTTATCATCTACAG 420  
 DB 404680 TTTTCTTGGCCCTGTTCCGAGCTGAATTATCCGCTACGCGATGTTTATCATCTACAG 404739  
 OY 421 AATTGCTGACCGTTTACAGGCCATCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 480  
 DB 404740 AATTGCTGACCGTTTACAGGCCATCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 404799  
 OY 481 TAAGAACTGAGCATTTTGTGCAATTTATGTAATGCTGTAACATGAGGAGGTATGC 540  
 DB 404800 TAAGAACTGAGCATTTTGTGCAATTTATGTAATGCTGTAACATGAGGAGGTATGC 404859  
 OY 541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTCAAAAATCCAGAGCAACACGTT 600  
 DB 404860 TATCTCTAATGTTCTGGGATTTGCTAATGAATTCAAAAATCCAGAGCAACACGTT 404919  
 OY 601 TCTACGCTCTTTTGGTATCGAGCTAATATTTCTTACTAGCTCTGCTGTCGTAATG 660  
 DB 404920 TCTACGCTCTTTTGGTATCGAGCTAATATTTCTTACTAGCTCTGCTGTCGTAATG 404979  
 OY 661 TTTGGGCTTCAAGATGAGAGCTTCCGTTTCTGAAGGTGATGCTTGGGCAATTTCTT 720  
 DB 404980 TTTGGGCTTCAAGATGAGAGCTTCCGTTTCTGAAGGTGATGCTTGGGCAATTTCTT 405039  
 OY 721 TACGCTTTTGAATGAGCTATGATGTAATCTGGAATGTTCTTATGGCCAGTACTGGT 780

```

|||||
Db 405040 TACGCTTTTGGATGCTATGACTATTGTAACGTGACTTGTCTTATAGGACAGTACTGCT 405099
QY 781 GGATCAATATAGAACGATTTAGACGATCCTGCTTATATATCAAGAAATGCAAAAG 840
Db 405100 GGATCATATAGAACGATTTAGACGATCCTGCTTATATATCAAGAAATGCAAAAG 405159
QY 841 GGAATAAAGGCTCTAAACCTAAATATGAAATGAAAGATAGCTTCTATCTTGGATGAT 900
Db 405160 GGAATAAAGGCTCTAAACCTAAATATGAAATGAAAGATAGCTTCTATCTTGGATGAT 405219
QY 901 CTCCTTATATCTTATATATCTCTCTGTTATATGCTTATATGATTTGATTTAACTTAA 960
Db 405220 CTCCTTATATCTTATATATCTCTCTGTTATATGCTTATATGATTTGATTTAACTTAA 405279
QY 961 TCGAAGTACTTGAATAAGCTGAAGTGAATGCAATATATATGATGATGATAGT 1020
Db 405280 TCGAAGTACTTGAATAAGCTGAAGTGAATGCAATATATGATGATGATAGT 405339
QY 1021 AGTTCAATGGGGAACCTTCTCTCTGAGCTGAGCTAGTATCCGTAATCATGCTATTTG 1080
Db 405340 AGTTCAATGGGGAACCTTCTCTCTGAGCTGAGCTAGTATCCGTAATCATGCTATTTG 405399
QY 1081 TTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 405400 TTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405459
QY 1141 TGGTCTCTCTAAGAGTATGTTTCTGCTCTGTTTCTGTTTGAACCAAGCTTCTG 1200
Db 405460 TGGTCTCTCTAAGAGTATGTTTCTGCTCTGTTTCTGTTTGAACCAAGCTTCTG 405519
QY 1201 GCGTGTGCTCTATGTTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 1260
Db 405520 GCGTGTGCTCTATGTTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 405579
QY 1261 AGAATATCTCTGCAAAATCCAAATATAGCTCTCTGCTCTGCTCTGCTCTGCTCTGCT 1320
Db 405580 AGAATATCTCTGCAAAATCCAAATATAGCTCTCTGCTCTGCTCTGCTCTGCTCTGCT 405639
QY 1321 ATATCCCTCTTGACCAAGAGCAAAAGTCAAAAGTAAAGGCTCTATGATGATGCTG 1380
Db 405640 ATATCCCTCTTGACCAAGAGCAAAAGTCAAAAGTAAAGGCTCTATGATGATGCTG 405699
QY 1381 CCGGCTTGGGAATATCAGAGAGAGCTTTATCCACAAAGGTTGCTCTGCTCTGCTG 1440
Db 405700 CCGGCTTGGGAATATCAGAGAGAGCTTTATCCACAAAGGTTGCTCTGCTCTGCTG 405759
QY 1441 GATTGGAGCTATGACCCCTTATCTTGAGTATCTCTTCTTCAATGCTATTTGCT 1500
Db 405760 GATTGGAGCTATGACCCCTTATCTTGAGTATCTCTTCTTCAATGCTATTTGCT 405819
QY 1501 TGGTTTCTGCACTAAGTTAAACAACTATTTAGCGAGTCTGCTCTTAAAGAACAG 1560
Db 405820 TGGTTTCTGCACTAAGTTAAACAACTATTTAGCGAGTCTGCTCTTAAAGAACAG 405879
QY 1561 AAGTGGCTCAAGAAATTCAGTCTGCTCTTCTTCAATAGGTTGCTCTTCTGCTGCT 1620
Db 405880 AAGTGGCTCAAGAAATTCAGTCTGCTCTTCTTCAATAGGTTGCTCTTCTGCTGCT 405939
QY 1621 GATCCCTACTGCTCTTT 1637
Db 405940 GATCCCTACTGCTCTTT 405956

```

RESULT 3  
AAC81914/C  
ID AAC81914 standard; DNA: 273254 BP.

AC AAC81914;  
XX  
XX 27-FEB-2001 (first entry)  
XX  
DE Chlamydia pneumoniae genome DNA.

```

XX Genome; diagnosis; vaccine; ds.
KW Chlamydia pneumoniae.
OS WO200027994-A2.
PN 18-MAY-2000.
XX 12-NOV-1999; 99WO-US26923.
PF 12-NOV-1999; 99US-0108279.
PR 08-APR-1999; 99US-0128606.
XX (REGC ) UNIV CALIFORNIA.
PA Stephens R, Mitchell W, Kalman S, Davis R;
XX WPI: 2000-376516/32.
DR
XX
XX Isolated nucleic acid for use in diagnostic and analytical methods
PT encodes genomic sequence of Chlamydia pneumoniae
PR
XX
XX Claim 2: Page 128-320; 320pp; English.
PS
XX
XX This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC pneumoniae protein (P1), given in the specification. The isolated nucleic
CC acid is useful for diagnostic and analytical methods, such as,
CC hybridization-based assays or amplification-based assays. The protein may
CC be used for diagnostic purposes, for their enzymatic or structural
CC activity, or as a vaccine. The invention also describes (1) a probe
CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC cassette comprising N1 under the transcriptional regulation of a
CC transcriptional termination region; (4) a cell comprising an expression
CC cassette of (3) as part of an extrachromosomal element or integrated into
CC the genome of a host cell, and the cellular progeny of the host cell;
CC (5) a method for producing a P1 comprising growing a cell of (4) where
CC the protein is expressed and isolating the protein free of other
CC proteins; (6) a purified polypeptide composition comprising at least 50
CC weight % of P1; and (7) a monoclonal antibody binding specifically to the
CC peptide of (6).
XX
SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;
Query Match 90.7%; Score 1484; DB 21; Length 273254;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAAATATAAAATATCAGATATGAAATATAAAGTATTTACAGAGGTTAAATATGACAAAAA 60
Db 212470 GAAATATAAAATATCAGATATGAAATATAAAGTATTTACAGAGGTTAAATATGACAAAAA 212411
QY 61 CCGAAGAAAACCTTTGGAAAATTCGCTCTTTCTTGTCGCGGATACATCTACGAGC 120
Db 212410 CCGAAGAAAACCTTTGGAAAATTCGCTCTTTCTTGTCGCGGATACATCTACGAGC 212351
QY 121 TAAAGAAAGTCTGCCAATGTTCTTAATGTTCTGATTAATCAATTAATACAGTGT 180
Db 212350 TAAAGAAAGTCTGCCAATGTTCTTAATGTTCTGATTAATCAATTAATACAGTGT 212291
QY 181 TACGCGATACAAAGACACTTTATTTGGAGAGCTCTGTTCTGTCGTCAGAGCAATAC 240
Db 212290 TACGCGATACAAAGACACTTTATTTGGAGAGCTCTGTTCTGTCGTCAGAGCAATAC 212231
QY 241 CTTTCATCAAGTTTGGCTGTTGTCCTGCTGCTATTAATCTTATATGCTATTTATGCA 300
Db 212230 CTTTCATCAAGTTTGGCTGTTGTCCTGCTGCTATTAATCTTATATGCTATTTATGCA 212171
QY 301 AGCTAAGTAATATTTAAGTAAGAGGCTTATTTATGCACTGGGACGCCCTTTTAA 360

```

```

Db 212170 AGCTAAGTAATATTTAAGTAAGCAGGCGTTATTTATGACATGGGAACGCCCTTTTAA 212111
OY 361 TTTCTTTGCCCTGTTCGCGACTAATTTATCCGCTACGCGATTTTACATCTCTACAG 420
Db 212110 TTTCTTTGCCCTGTTCGCGACTAATTTATCCGCTACGCGATTTTACATCTCTACAG 212051
OY 421 AATTTGCTGACCGTTTACAGGCGATCTACCTCAGAGATGCTAGAGACTCGTTGCCATCT 480
Db 212050 AATTTGCTGACCGTTTACAGGCGATCTACCTCAGAGATGCTAGAGACTCGTTGCCATCT 211991
OY 481 TAAGAACTGCGACTTTGCTGATTTTATGTACTGCTGAGACTATGAGGGAAGCGTCATCG 540
Db 211990 TAAGAACTGCGACTTTGCTGATTTTATGTACTGCTGAGACTATGAGGGAAGCGTCATCG 211931
OY 541 TATCCTATGTTGCGGGAATTTGCTAATGAATACAAAATTCACGAGCAAGCAAGCGCT 600
Db 211930 TATCCTATGTTGCGGGAATTTGCTAATGAATACAAAATTCACGAGCAAGCAAGCGCT 211871
OY 601 TCTACGCTCTTTTCGCTACGAGCTAATATTTCTTTACTAGCTCTGCTGCGCAATTCG 660
Db 211870 TCTACGCTCTTTTCGCTACGAGCTAATATTTCTTTACTAGCTCTGCTGCGCAATTCG 211811
OY 661 TTTGGGCTTCAAGTTGAGAGCTCCGTTTCTGAAGGTGTAGATCCCTGGGGAATTTCTT 720
Db 211810 TTTGGGCTTCAAGTTGAGAGCTCCGTTTCTGAAGGTGTAGATCCCTGGGGAATTTCTT 211751
OY 721 TACGCTCTTTGATGGCTATGACTATTTGATCTGAGACTTGTCTTATGGCAATTTCTGT 780
Db 211750 TACGCTCTTTGATGGCTATGACTATTTGATCTGAGACTTGTCTTATGGCAATTTCTGT 211691
OY 781 GGATCAATTAAGAACGTAATTTGACGAGTCCGCTTCTATATACCAAGAAATGACAAAGG 840
Db 211690 GGATCAATTAAGAACGTAATTTGACGAGTCCGCTTCTATATACCAAGAAATGACAAAGG 211631
OY 841 GGAAGAAAGGTGCTAACCCTAAATGAATGAAGAGATGCTTCTATCTTATGATAGAT 900
Db 211630 GGAAGAAAGGTGCTAACCCTAAATGAATGAAGAGATGCTTCTATCTTATGATAGAT 211571
OY 901 CTCCTTATATCTTTTATTAATCTCTGCTGATGCGATGCGATGCTTATGCAATTAATCA 960
Db 211570 CTCCTTATATCTTTTATTAATCTCTGCTGATGCGATGCGATGCTTATGCAATTAATCA 211511
OY 961 TCGAAGTGAATTTGGAAGATCAGTGAATGCAATATCTCAATATGCAATGCTATAGTG 1020
Db 211510 TCGAAGTGAATTTGGAAGATCAGTGAATGCAATATCTCAATATGCAATGCTATAGTG 211451
OY 1021 AGTTCATGGGGAATCTTCTCTCTGAGCTGCGGTAGTATCCGTAATCATGCTATTGG 1080
Db 211450 AGTTCATGGGGAATCTTCTCTCTGAGCTGCGGTAGTATCCGTAATCATGCTATTGG 211391
OY 1081 TTGGTGTGTAACGTCATGCTTAATTTGATGCTTAACGCGACCCCTAGTCACTCCTGTCA 1140
Db 211390 TTGGTGTGTAACGTCATGCTTAATTTGATGCTTAACGCGACCCCTAGTCACTCCTGTCA 211331
OY 1141 TGGTTCCTCTTAACAGGTATCGTTTCTTCGCTCTGTTATCTTTGAAACCAAGCTTCG 1200
Db 211330 TGGTTCCTCTTAACAGGTATCGTTTCTTCGCTCTGTTATCTTTGAAACCAAGCTTCG 211271
OY 1201 GCGTGTGCTATGTTGCGTACACTCTCTCATCTAGCTGTGTTGTGCGAGGCTATAC 1260
Db 211270 GCGTGTGCTATGTTGCGTACACTCTCTCATCTAGCTGTGTTGTGCGAGGCTATAC 211211
OY 1261 AGAATATTTCTTTCGAAATTCACAAATATGCTCTTTGAGTCAACTAAAGAAATGGCCT 1320
Db 211210 AGAATATTTCTTTCGAAATTCACAAATATGCTCTTTGAGTCAACTAAAGAAATGGCCT 211151
OY 1321 ATATCCCTCTTTCGAAAGAGCAAAAGTCAAGGCTCTATGATGTATGATGCTGGA 1380
Db 211150 ATATCCCTCTTTCGAAAGAGCAAAAGTCAAGGCTCTATGATGTATGATGCTGGA 211091
OY 1381 CCGGCTTTCGAAAGATCAGAGAGCTTTATATCAACAAGGTTGCTGTTATCTGTGGA 1440
Db 211090 CCGGCTTTCGAAAGATCAGAGAGCTTTATATCAACAAGGTTGCTGTTATCTGTGGA 211031

```

```

OY 1441 GATTTGAGCTATGACCCCTTATCTTGCAGTATCTCTTTCAATGCTATTGGT 1500
Db 211030 GATTTGAGCTATGACCCCTTATCTTGCAGTATCTCTTTCAATGCTATTGGT 210971
OY 1501 TGGTTCCTGCAACTAAGTTAAACAACATATCTTATGCGCAGTCTGCTTTAAACAAG 1560
Db 210970 TGGTTCCTGCAACTAAGTTAAACAACATATCTTATGCGCAGTCTGCTTTAAACAAG 210911
OY 1561 AAGTGGCTTCAGACATTCAGCTCTGCTCTTTCATPAGAGTTGCTTCTTACTTGT 1620
Db 210910 AAGTGGCTTCAGACATTCAGCTCTGCTCTTTCATPAGAGTTGCTTCTTACTTGT 210851
OY 1621 GATCCCTACCTGCTT 1637
Db 210850 GATCCCTACCTGCTT 210834

```

## RESULT 4

AAZ01425 standard; DNA; 1038602 BP.

AAZ01425;

07-OCT-1999 (first entry)

Complete genome sequence of Chlamydia trachomatis.

Vaccine: eye disease; conventional trachoma; nonendemic trachoma;  
 Paratrachoma; inclusion conjunctivitis; genital disease; perithenitis;  
 nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 Bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.

Chlamydia trachomatis.

W09928475-A2.

10-JUN-1999.

27-NOV-1998; 98WO-1B01939.

04-NOV-1998; 98US-0107077.

28-NOV-1997; 97FR-0015041.

17-DEC-1997; 97FR-0016034.

(GBST ) GENSET.

Griffais R;

Genome sequence of Chlamydia trachomatis

Claim 1: Page 373-656; 1755pp; English.

The present sequence represents the complete genome of Chlamydia trachomatis. Open reading frames (ORFs) of the genome encode polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perithenitis, Bartholinitis; CC lymphoparatomy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.

Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other:

Query Match 2.4%; Score 39; DB 20; Length 1038602;

Best Local Similarity 100.0%; Pred. No. 5.7e-09;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 AAGAAAGTCTGCAATGTTCTTAATGTTCTTCTATT 161  
 |||||||  
 DB 392479 AAGAAAGTCTGCAATGTTCTTAATGTTCTTCTATT 392517

## RESULT 5

AAA30923  
 ID AAA30923 standard; DNA; 43 BP.

AC AAA30923;

DT 22-SEP-2000 (first entry)

DE PCR primer for C. pneumoniae ATP/ADP translocase coding sequence.

KM ATP/ADP translocase; Chlamydia infection; diagnosis; therapy;  
 KM PCR primer; ss.

OS Chlamydia pneumoniae.

PN WO200039157-A1.

PD 06-JUL-2000.

PF 22-DEC-1999; 99WO-CA01224.

PR 28-DEC-1998; 98US-0114060.

PR 12-MAR-1999; 99US-0123967.

PR 30-JUN-1999; 99US-0141271.

PA (CONN-) CONNAUGHT LAB LTD.

PI Murdin AD, Oomen RP, Wang J, Dunn P;

DR WPI; 2000-452368/39.

PT Novel Chlamydia polynucleotides and polypeptides, useful for diagnosis,  
 prevention and treatment of Chlamydia infection in mammals -

PS Claim 32; Page 48; 81pp; English.

CC This sequence represents a PCR primer used to isolate DNA encoding  
 the Chlamydia pneumoniae ATP/ADP translocase of the invention. The  
 CC protein, DNA encoding it, or a vaccine containing the DNA or protein, are  
 CC useful for diagnosing, preventing or treating Chlamydia infection. The  
 CC sequences can also be used in a method for the detection of Chlamydia  
 CC infection. Primers or probes derived from the DNA sequence are useful in  
 CC diagnostic tests for detecting Chlamydia infection.

SO Sequence 43 BP; 20 A; 12 C; 8 G; 3 T; 0 other;

Query Match 1.4%; Score 23; DB 21; Length 43;

Best Local Similarity 100.0%; Pred. No. 0.57; Mismatches 0; Indels 0; Gaps 0;

OY 51 ATGACAAAACCGAAGAAAAC 73  
 |||||||

DB 21 ATGACAAAACCGAAGAAAAC 43

## RESULT 6

AAAL23777  
 ID AAAL23777 standard; cDNA; 786 BP.

AC AAAL23777;

DT 07-DEC-2001 (first entry)

DE Human breast cancer expressed polynucleotide 16234.

KM Human; breast cancer; cell marker; cytostatic; ss.  
 KM Human; breast cancer; cell marker; cytostatic; ss.

OS Homo sapiens.

PN WO200151628-A2.

PD 19-JUL-2001.

PF 10-JAN-2001; 2001WO-US00798.

PR 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

PR 24-MAR-2000; 2000US-0192099.

PR 29-MAR-2000; 2000US-0193480.

PR 15-MAY-2000; 2000US-0205230.

PR 09-JUN-2000; 2000US-0211315.

PR 25-JUL-2000; 2000US-0220534.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Lillie J, Xu Y, Wang Y, Steinmann K;

DR WPI; 2001-451856/48.

PT New peptide useful as a marker for the diagnosis of breast cancer -

PS Claim 1; Page 2966; 3695pp; English.

CC The invention relates to human breast cancer expressed polynucleotides  
 CC (AAI07544-AI26789) and methods of assessing whether a patient is  
 CC afflicted with breast cancer by examining the correlation between the  
 CC expression of certain markers and the cancerous state of breast cells.  
 CC The polynucleotides and encoded polypeptides are potential markers for  
 CC detecting, diagnosing, monitoring, characterizing treating and  
 CC potentially preventing breast cancer. The polynucleotides and encoded  
 CC polypeptides are also useful for isolating compounds with cytostatic  
 CC activity.

SO Sequence 786 BP; 201 A; 177 C; 169 G; 238 T; 1 other;

Query Match 1.3%; Score 22; DB 22; Length 786;

Best Local Similarity 100.0%; Pred. No. 1.7; Mismatches 0; Indels 0; Gaps 0;

OY 908 TATTTCTTTTATTACTCTCTTG 929  
 |||||||

DB 485 TATTTCTTTTATTACTCTCTTG 506

## RESULT 7

AAS30049  
 ID AAS30049 standard; DNA; 2139 BP.

AC AAS30049;

DT 21-NOV-2001 (first entry)

DE Human lung antigen genomic DNA #119.

KM Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;

KM chicken; sheep; immunosuppressive; antitubercular; vasotropic;

KM antipneumatic; antiproliferative; cytostatic; cardiant; neuroprotective;

KM cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;

KM ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;

KM hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;

KM cerebrovascular disorder; nervous system disorder; bacterial infection;

KM fungal infection; viral infection; ocular disorder; endocrine disorder;

KM gastrointestinal disorder; renal disorder; respiratory disorder;

KM wound healing; skin aging; organ transplantation; food preservative;

KM tissue regeneration; anti-infertility; food additive.

OS Homo sapiens.

PN WO200155303-A2.

PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01301.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-457723/49.  
DR

XX Isolated polypeptide for treating, preventing and/or prognosing  
PT respiratory disorders related to the lung including lung cancers and  
PT also for testing and detection e.g. diagnosis -  
PS Claim 1; SEQ ID NO 313; 507pp; English.  
XX  
CC Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode  
CC the lung antigen polypeptides of the invention. Lung antigen polypeptides  
CC and their associated polynucleotides are useful in the diagnosis,  
CC treatment and prevention of various types of disorders in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A  
CC pathological condition can be determined by detecting the presence or  
CC absence of a mutation in a lung antigen polynucleotide. The treatable  
CC disorders include autoimmune diseases such as rheumatoid arthritis,  
CC hyperproliferative disorders such as neoplasms of the breast or liver,  
CC cardiovascular disorders such as cardiac arrest, cerebrovascular  
CC disorders such as cerebral ischaemia, nervous system disorders such as  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
CC ocular disorders such as corneal infection, endocrine disorders such as  
CC premature labour and infertility, gastrointestinal disorders such as  
CC Crohn's disease, renal disorders such as glomerulonephritis and  
CC respiratory disorders such as asthma and pleurisy. The polypeptides can  
CC also be used to aid wound healing, to prevent skin aging due to sunburn,  
CC to maintain organs before transplantation, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 1.3%; Score 21; DB 22; Length 2139;  
Best Local Similarity 100.0%; Pred. No. 5.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 700 TAGATCCTTGGGGAATTCTT 720  
|||||  
Db 31 TAGATCCTTGGGGAATTCTT 51

RESULT 8  
AAS30048  
ID AAS30048 standard; DNA; 4673 BP.  
XX  
AC AAS30048;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human lung antigen genomic DNA #118.  
XX  
XX Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;  
KM chicken; sheep; immunosuppressive; antiarthritic; vasotropic;  
KM antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KM cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
KM ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;  
KM hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KM cerebrovascular disorder; nervous system disorder; bacterial infection;  
KM fungal infection; viral infection; ocular disorder; endocrine disorder;  
KM gastrointestinal disorder; renal disorder; respiratory disorder;  
KM wound healing; skin aging; organ transplantation; food preservative;  
KM tissue regeneration; anti-infertility; food additive.  
XX  
OS Homo sapiens.  
XX  
XX WO200155303-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01301.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR  
XX 04-FEB-2000; 2000US-0180628.  
PR  
XX 24-FEB-2000; 2000US-0184664.  
PR

PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227059.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229309.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.



throughput method for detecting altered expression of one or more polynucleotides in a sample. The polynucleotides can be used in the diagnosis of disorders associated with foam cell development such as atherosclerosis, cerebral stroke, and cardiovascular disorders such as coronary artery disease. The polynucleotide sequences can also be used as PCR primers and probes. The polynucleotides of the invention are also useful in gene therapy. AAS94746-AAS95021 represent the human polynucleotide sequences of the invention which are differentially expressed during foam cell differentiation.

Sequence 5032 BP; 1473 A; 1026 C; 1055 G; 1477 T; 1 other;

Query Match 1.3%; Score 21; DB 24; Length 5032;  
Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 TAGATCCTTGGGAATTTCTT 720  
|||||  
DB 2927 TAGATCCTTGGGAATTTCTT 2947

RESULT 10

AA083529  
ID AA083529 standard; DNA; 8220 BP.

XX AA083529;

DT 22-SEP-1995 (first entry)

XX P. falciparum Proj3 gene.

XX Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;

XX vaccine; ss.

OS Plasmodium falciparum.

XX W09507353-A.

XX 16-MAR-1995.

XX 07-SEP-1994; 94WO-US10230.

XX 10-SEP-1993; 93US-0119677.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

XX Wellens TE;

XX WPI: 1995-123427/16.

XX P-PSDB; AAR70236.

XX New erythrocyte binding domain polypeptide(s) - isolated from

XX Plasmodium binding proteins, used in diagnosis, treatment and

XX prevention of malaria

XX Disclosure: Page 61-65; 81pp; English.

XX Erythrocyte binding ligand (EBL) family genes were cloned from

XX P. falciparum chromosome 7 subsegment libraries constructed during

XX genetic studies of the chloroquine resistance locus. The 4 genes,

XX EBL-e1 (AA083526), E31a (AA083527), EBL-e2 (AA083528) and Proj3

XX (AA083529), encode the proteins given in AAR70233-36, respectively. The

XX binding domains of such proteins can be expressed e.g. in E. coli,

XX yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected

XX cells, and provide protection against P. falciparum.

XX Sequence 8220 BP; 3194 A; 1249 C; 1777 G; 2000 T; 0 other;

XX Query Match 1.3%; Score 21; DB 16; Length 8220;

XX Best Local Similarity 100.0%; Pred. No. 5.2;

XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAACCTTTGGGAAAT 84  
|||||  
DB 5046 AAGAAAACCTTTGGGAAAT 5066

RESULT 11

AAAT72897  
ID AAT72897 standard; cDNA; 8220 BP.

XX AAT72897;

DT 07-OCT-1997 (first entry)

XX Plasmodium Proj3 gene.

XX DBL gene family; SABB; stialic acid binding proteol; vaccine; therapy;

XX Duffy binding like gene; Duffy antigen binding protein; erythrocyte;

XX DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;

XX Plasmodium; ss.

XX Plasmodium falciparum.

XX Key Location/Qualifiers

XX CDS 2..8113

XX /\*tag= a

XX /note= "encodes residues 8-2710 of AAW22482"

XX W09640766-A2.

XX 19-DEC-1996.

XX 07-JUN-1996; 96WO-US09508.

XX 07-JUN-1995; 95US-0487826.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

XX Wellens TE;

XX WPI: 1997-052231/05.

XX P-PSDB; AAW22482.

XX New malaria vaccines - contains cysteine-rich DBL family protein

XX binding domains homologous domains of the Duffy and stialic acid

XX binding proteins

XX Disclosure: Page 48-50; 96pp; English.

XX This sequence represents the Proj3 gene of Plasmodium. Proj3 belongs to

XX the Duffy binding like (DBL) family of genes which have homology to the

XX Duffy antigen binding protein (DABP) and stialic acid binding protein

XX (SABB) conserved regions (see AAT72889 and AAT72888 respectively). The

XX var family of genes modulate cytoadherence and antigenic variation of

XX Plasmodium infected erythrocytes. SABB and the Duffy antigen binding

XX supernatant after infected erythrocytes release merozoites. DABP and

XX SABB mediate the binding of merozoites and schizonts to the erythrocyte

XX surface. These proteins are necessary for erythrocyte invasion by the

XX parasite. This sequence can be used in the compositions of the

XX invention. The compositions are for the treatment and prevention of

XX malaria, and comprise either a nucleotide sequence or encoded polypeptide

XX of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a

XX family of genes having homology with conserved regions of DABP and SABB.

XX The compositions are used for the treatment and prevention of malaria.

XX They are also used in the preparation of vaccines for inducing a

XX protective immune response in a mammal to Plasmodium merozoites

XX (especially Plasmodium falciparum or Plasmodium vivax).

XX Sequence 8220 BP; 3194 A; 1249 C; 1777 G; 2000 T; 0 other;

XX Query Match 1.3%; Score 21; DB 18; Length 8220;

XX Best Local Similarity 100.0%; Pred. No. 5.2;



Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAAACCTTTGGAAAT 84  
 |||||||||||||||||||  
 Db 5046 AAGAAAAACCTTTGGAAAT 5066

RESULT 12  
 AA298286  
 ID AA298286 standard; DNA; 8220 BP.  
 XX  
 AC AA298286;  
 XX  
 DT 13-JUN-2000 (first entry)  
 XX  
 DE P. falciparum Proj3 binding domain polypeptide encoding DNA.  
 XX  
 KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;  
 KW DABP; Stalic Acid Binding Protein; SABB; malaria; vaccine; immunisation;  
 KW protozoacide; ds.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 PN US5993827-A.  
 XX  
 PD 30-NOV-1999.  
 XX  
 PF 07-JUN-1995; 95US-0487826.  
 XX  
 PR 10-SEP-1993; 93US-0119677.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;  
 XX  
 DR WPI; 2000-194198/17.  
 XX  
 DR P-PSDB; AAT77904.  
 XX  
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium  
 PT falciparum erythrocyte binding proteins useful for vaccinating against  
 PT malaria -  
 XX  
 PS Disclosure; Columns 71-80; 93pp; English.  
 XX  
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBL  
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially  
 CC identical to the Duffy Antigen Binding Protein (DABP) and Stalic Acid  
 CC Binding Protein (SABP), which are soluble proteins that appear in the  
 CC culture supernatant after erythrocytes infected with malaria release  
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the  
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum. Duffy  
 CC and stalic acid receptors on erythrocytes. The ebl-1 polypeptides may be  
 CC used to vaccinate against malaria, especially caused by P. falciparum.  
 CC Immunization with the polypeptide provides effective protection against  
 CC malaria. The present sequence represents the Proj3 binding domain  
 CC polypeptide encoding DNA.  
 XX  
 SQ Sequence 8220 BP; 3194 A; 1249 C; 1777 G; 2000 T; 0 other;

Query Match 1.3%; Score 21; DB 21; Length 8220;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAAACCTTTGGAAAT 84  
 |||||||||||||||||||  
 Db 5046 AAGAAAAACCTTTGGAAAT 5066

RESULT 13  
 AAT72882  
 ID AAT72882 standard; CDNA; 19124 BP.  
 XX  
 AC AAT72882;  
 XX

XX  
 DT 12-SEP-1997 (first entry)  
 XX  
 DE Plasmodium var-7 gene.  
 XX  
 KW DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;  
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;  
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;  
 KW Plasmodium; ss.  
 XX  
 OS Plasmodium vivax.  
 OS Plasmodium falciparum.  
 XX  
 FH Key location/Qualifiers  
 XX  
 FT exon 7317..15139  
 FT /\*tag= a  
 FT /number= 1  
 FT 15140..16205  
 FT /\*tag= b  
 FT /number= 1  
 FT 16206..17552  
 FT exon 16206..17552  
 FT /\*tag= c  
 FT /number= 2  
 FT /note= "no stop codon given"  
 XX  
 PN WO9640766-A2.  
 XX  
 PD 19-DEC-1996.  
 XX  
 PF 07-JUN-1996; 96WO-US09508.  
 XX  
 PR 07-JUN-1995; 95US-0487826.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;  
 XX  
 DR Wellens TE;  
 XX  
 DR WPI; 1997-052231/05.  
 XX  
 DR P-PSDB; AAW2475.  
 XX  
 PT New malaria vaccines - contains cysteine-rich DBL family protein  
 PT binding domains homologous domains of the Duffy and stalic acid  
 PT binding proteins  
 XX  
 PS Claim 4; Page 56-61; 96pp; English.  
 XX  
 CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to  
 CC the Duffy binding like (DBL) family of genes which have homology to the  
 CC Duffy antigen binding protein (DABP) and stalic acid binding protein  
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The  
 CC var family of genes modulate cytoadherence and antigenic variation of  
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding  
 CC protein (DABP) are soluble proteins that appear in the culture  
 CC supernatant after infected erythrocytes release merozoites. DABP and  
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte  
 CC surface. These proteins are necessary for erythrocyte invasion by the  
 CC parasite. This sequence can be used in the compositions of the  
 CC invention. The compositions can be used for the treatment and prevention of  
 CC malaria. The sequence either a nucleotide sequence or encoded polypeptide  
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family.  
 CC family of genes having homology with conserved regions of DABP and SABP.  
 CC The compositions are used for the treatment and prevention of malaria.  
 CC They are also used in the preparation of vaccines for inducing a  
 CC protective immune response in a mammal to Plasmodium merozoites  
 CC (especially Plasmodium falciparum or Plasmodium vivax).  
 XX  
 SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 1.3%; Score 21; DB 18; Length 19124;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAACCTTTGGAAAT 84  
 ID 12366 AAGAAAACCTTTGGAAAT 12386  
 RESULT 14  
 AA298287  
 ID AA298287 standard; DNA: 19124 BP.  
 XX  
 AC AA298287;  
 XX  
 DT 13-JUN-2000 (first entry)  
 XX  
 DE Plasmodium var-7 polypeptide encoding DNA.  
 XX  
 KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;  
 KM DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;  
 KM protozoacide; var-7; ds.  
 XX  
 OS Plasmodium sp.  
 XX  
 PN US593827-A.  
 PD 30-NOV-1999.  
 XX  
 PF 07-JUN-1995; 95US-0487826.  
 XX  
 PR 10-SEP-1993; 93US-0119677.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;  
 DR WPI: 2000-194198/17.  
 DR P-PSDB; AAY77904.  
 XX  
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium  
 PT falciparum erythrocyte binding proteins useful for vaccinating against  
 PT malaria -  
 XX  
 PS Disclosure; Columns 91-108; 93pp; English.  
 XX  
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBL  
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially  
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid  
 CC Binding Protein (SABP), which are soluble proteins that appear in the  
 CC culture supernatant after erythrocytes infected with malaria release  
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the  
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy  
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be  
 CC used to vaccinate against malaria, especially caused by P. falciparum.  
 CC Immunization with the polypeptide provides effective protection against  
 CC malaria. The present sequence represents the var-7 polypeptide encoding  
 CC DNA.  
 CC  
 SO Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;  
 Query Match 1.3%; Score 21; DB 21; Length 19124;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 64 AAGAAAACCTTTGGAAAT 84  
 ID 12366 AAGAAAACCTTTGGAAAT 12386  
 RESULT 15  
 AAX95295/C  
 ID AAX95295 standard; DNA: 20 BP.  
 AC AAX95295;  
 XX  
 DT 13-SEP-1999 (first entry)

XX  
 DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
 XX  
 KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KM vaccine; neutralising epitope; PCR primer; ss.  
 XX  
 OS Synthetic.  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-1B01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Grifffais R;  
 XX  
 DR WPI: 1999-357842/30.  
 XX  
 XX Genome sequence of Chlamydia pneumoniae  
 PT  
 PS Page 1737; Disclosure: 1912pp; English.  
 XX  
 CC AAX91991-X97517 represent PCR primers used to amplify open reading  
 CC frames and other nucleic acid sequences from the genome of  
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
 CC disease such as pneumonia and bronchitis and is thought to be a  
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
 CC containing C. pneumoniae nucleotides sequences can also be used as  
 CC immunogenic compositions, especially where the vector directs the  
 CC expression of a neutralising epitope of C. pneumoniae.  
 CC  
 SO Sequence 20 BP; 7 A; 4 C; 5 G; 4 T; 0 other;  
 Query Match 1.2%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 72 CCTTTGGAAATGCGCTC 91  
 ID 20 CCTTTGGAAATGCGCTC 1  
 DB  
 RESULT 16  
 AAX95297/C  
 ID AAX95297 standard; DNA: 20 BP.  
 AC AAX95297;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; PCR primer; ss.  
 XX  
 OS Synthetic.  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO927105-A2.  
 XX  
 PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.  
XX  
PA (GEST ) GENSET.  
XX  
PI Griffais R:  
XX WPI; 1999-357842/30.  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1737; Disclosure: 1912pp; English.  
XX  
CC AAX91991-X97517 represent PCR primers used to amplify open reading  
CC frames and other nucleic acid sequences from the genome of  
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
CC disease such as pneumonia and bronchitis and is thought to be a  
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
CC containing C. pneumoniae nucleotides sequences can also be used as  
CC immunogenic compositions, especially where the vector directs the  
CC expression of a neutralising epitope of C. pneumoniae.  
XX  
SQ Sequence 20 BP; 6 A; 7 C; 3 G; 4 T; 0 other;  
XX  
Query Match 1.2%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1018 GTGAGTTCATGGGGAACCTTC 1037  
DB 20 GTGAGTTCATGGGGAACCTTC 1  
XX  
RESULT 17  
AAX95298/C  
ID AAX95298 standard; DNA; 20 BP.  
XX  
AC AAX95298;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope; PCR primer; ss.  
XX  
XX Synthetic.  
OS Chlamydia pneumoniae.  
XX  
PN WO9927105-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.  
XX  
PA (GEST ) GENSET.  
XX  
PI Griffais R:  
XX WPI; 1999-357842/30.  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1737; Disclosure: 1912pp; English.

XX  
CC AAX91991-X97517 represent PCR primers used to amplify open reading  
CC frames and other nucleic acid sequences from the genome of  
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
CC disease such as pneumonia and bronchitis and is thought to be a  
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
CC containing C. pneumoniae nucleotides sequences can also be used as  
CC immunogenic compositions, especially where the vector directs the  
CC expression of a neutralising epitope of C. pneumoniae.  
XX  
SQ Sequence 20 BP; 9 A; 5 C; 5 G; 1 T; 0 other;  
XX  
Query Match 1.2%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 83 ATTCGCTCTTCTTGTGGC 102  
DB 20 ATTCGCTCTTCTTGTGGC 1  
XX  
RESULT 18  
AAX92828  
ID AAX92828 standard; DNA; 20 BP.  
XX  
AC AAX92828;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope; PCR primer; ss.  
XX  
XX Synthetic.  
OS Chlamydia pneumoniae.  
XX  
PN WO9927105-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.  
XX  
PA (GEST ) GENSET.  
XX  
PI Griffais R:  
XX WPI; 1999-357842/30.  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1542; Disclosure: 1912pp; English.  
XX  
CC AAX91991-X97517 represent PCR primers used to amplify open reading  
CC frames and other nucleic acid sequences from the genome of  
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
CC disease such as pneumonia and bronchitis and is thought to be a  
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
CC containing C. pneumoniae nucleotides sequences can also be used as  
CC immunogenic compositions, especially where the vector directs the  
CC expression of a neutralising epitope of C. pneumoniae.  
XX  
SQ Sequence 20 BP; 5 A; 3 C; 7 G; 5 T; 0 other;

Query Match 1.2%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1244 GGTTGCGAGCTATACAGA 1263  
DB 1 GGTTGCGAGCTATACAGA 20  
|||||

RESULT 19  
AAx92823  
ID AAX92823 standard; DNA; 20 BP.  
XX  
AC AAX92823;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope; PCR primer; ss.  
XX  
OS Synthetic.  
OS Chlamydia pneumoniae.  
XX  
PN WO9927105-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.  
XX  
PA (GEST ) GENSET.  
XX  
PI Grifffals R;  
PI  
DR WPI; 1999-357842/30.  
XX  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1541; Disclosure; 1912pp; English.  
XX  
CC AAX91991-X97517 represent PCR primers used to amplify open reading  
CC frames and other nucleic acid sequences from the genome of  
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
CC disease such as pneumonia and bronchitis and is thought to be a  
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
CC containing C. pneumoniae nucleotides sequences can also be used as  
CC immunogenic compositions, especially where the vector directs the  
CC expression of a neutralising epitope of C. pneumoniae.  
XX  
SQ Sequence 20 BP; 4 A; 6 C; 4 G; 6 T; 0 other;

Query Match 1.2%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1446 GGAGCTATGACCCCTTATCT 1465  
DB 1 GGAGCTATGACCCCTTATCT 20  
|||||

RESULT 20  
AAx92826  
ID AAX92826 standard; DNA; 20 BP.  
XX

AC AAX92826;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope; PCR primer; ss.  
XX  
OS Synthetic.  
OS Chlamydia pneumoniae.  
XX  
PN WO9927105-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.  
XX  
PA (GEST ) GENSET.  
XX  
PI Grifffals R;  
PI  
DR WPI; 1999-357842/30.  
XX  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1542; Disclosure; 1912pp; English.  
XX  
CC AAX91991-X97517 represent PCR primers used to amplify open reading  
CC frames and other nucleic acid sequences from the genome of  
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
CC disease such as pneumonia and bronchitis and is thought to be a  
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
CC containing C. pneumoniae nucleotides sequences can also be used as  
CC immunogenic compositions, especially where the vector directs the  
CC expression of a neutralising epitope of C. pneumoniae.  
XX  
SQ Sequence 20 BP; 2 A; 4 C; 5 G; 9 T; 0 other;

Query Match 1.2%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 924 CTTGTTATGCGCTATGG 943  
DB 1 CTTGTTATGCGCTATGG 20  
|||||

RESULT 21  
AAF61373  
ID AAF61373 standard; cDNA; 1770 BP.  
XX  
AC AAF61373;  
XX  
DT 25-MAY-2001 (first entry)  
XX  
DE A. thaliana chloroplast ATP/ADP translocator cDNA.  
XX  
KW ATP/ADP translocator; chloroplast; plant; adenosine triphosphate;  
KW adenosine diphosphate; animal feed; energy; biosynthesis;  
KW nutritional value; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200120009-A1.  
XX

PD 22-MAR-2001.  
XX  
PF 05-AUG-2000; 2000MO-EP07625.  
XX  
PR 15-SEP-1999; 99DE-1044212.  
XX  
PA (BADI ) BASF AG.  
XX  
PI Reindl A, Geigenberger PL, Neuhaus H, Graeve-Kampfenkel K;  
PI Moehlmann T, Tjaden J;  
XX  
DR WPI: 2001-244814/25.  
XX  
PT Transformed plant with an altered content of essential amino acids and  
PT having improved nutritional quality, for use as an animal feed,  
PT comprises a modification in the ATP/ADP translocator gene  
XX  
XX  
PS Claim 6; Fig 1; 32pp; German.  
XX  
CC This invention describes a novel transformed plant (A), and its  
CC descendants, in which the regulatory sequences and/or the gene copy  
CC number of an ATP/ADP (adenosine tri-/di-phosphate) translocator gene  
CC (1) has been altered so that, compared with the wild-type plant, it  
CC has a different content of one or more amino acids (aa). The invention  
CC also describes (1) (1), for use in plants, having an approximately 1.7  
CC kb sequence (S1), given in the specification, from Arabidopsis thaliana,  
CC EMBL 249227; (2) a genetic construct (GC) containing the (1) and operably  
CC linked regulatory sequences; (3) a vector containing the (1) or GC; and  
CC (4) seeds, tissues, cells and replicative materials of (A). (A) are  
CC useful for human or animal feeding, also (including their cells, tissues  
CC and extracts) in agriculture, the animal feed and pharmaceutical  
CC industries, and the health service. Altering the activity of (1)  
CC increases the amount of ATP in chloroplasts and the amount of energy  
CC available for biosynthesis. Specifically (A) have increased contents of  
CC one or more essential aa, so are of greater nutritional value than  
CC wild-type plants.  
XX  
SQ Sequence 1770 BP; 441 A; 383 C; 431 G; 515 T; 0 other;  
  
Query Match 1.2%; Score 20; DB 22; Length 1770;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1308 AAAGAAATGGCCATATGCC 1327  
Db 1489 AAAGAAATGGCCATATGCC 1508  
  
RESULT 22  
AAK77984/C  
ID AAK77984 standard; DNA; 2287 BP.  
XX  
AC AAK77984;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32796.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX  
 DR WPI; 2001-483426/52.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 XX  
 PS Disclosure; SEQ ID NO 32796; 3071pp + Sequence Listing; English.  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (1)

CC amino acid sequences given in AAM82170 to AAM91921. (1) have cytosolic  
 CC activity, and can be used in gene therapy and vaccine production. (1)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (1) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (1) by expressing inactive proteins or to  
 CC supplement the patient's own production of (1). Additionally, (1)  
 CC polynucleotides may be used to produce the secreted (1), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (1) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK67694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 SO Sequence 2287 BP; 843 A; 386 C; 428 G; 630 T; 0 other;  
 Query Match 1.2%; Score 20; DB 22; Length 2287;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 714 ATTCTTTACGCTTTTGAT 733  
 Db 645 ATTCTTTACGCTTTTGAT 626  
 RESULT 23  
 AAM82169  
 ID AAM82169 standard; DNA; 33780 BP.  
 AC AAM82169;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Nucleotide sequence of disrupted in schizophrenia 1 (DIS1) intron 8.  
 XX  
 KW Disrupted in schizophrenia 1 gene; DIS1 gene; schizophrenia;  
 KW psychiatric disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200140301-A2.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PE 28-NOV-2000; 2000WO-EP11915.  
 XX  
 PF 01-DEC-1999; 99EP-0309667.  
 XX  
 PA (ALKU) AKZO NOBEL NV.  
 PA (MED1-) MEDICAL RES COUNCIL.  
 PA (UYED-) UNIV EDINBURGH.  
 XX  
 PI Porteous D, Millar K, Blackwood D;  
 XX  
 DR WPI; 2001-374796/39.  
 XX  
 PT Novel isolated polynucleotide which surrounds a breakpoint on  
 PT chromosome 1 involved in a balanced t(1;11)(q42.1;q14.3)  
 PT translocation, and its encoded proteins, useful as medicament for  
 PT treating psychiatric disorders -  
 XX  
 PS Disclosure; Page 39-50; 51pp; English.  
 XX  
 CC The present sequence represents intron 8 of a human disrupted in  
 CC schizophrenia 1 (DIS1) gene. DIS1 is disrupted by a (1;11)(q42.1;q14.3)  
 CC translocation which segregates with schizophrenia. The DIS1 gene consists  
 CC of 13 exons. Exon 11 contains an alternative splice site which removed 66  
 CC nucleotides from the open reading frame. The final intron of DIS1 belongs  
 CC to the rare A/C class of introns. DIS1 polypeptide fragments and  
 CC antibodies are useful for in vitro diagnosis of a psychiatric disorder.

CC DIS1 protein is useful in a screening assay for identifying new drugs for  
CC treating psychiatric disorders. DIS1 polypeptides and polynucleotides are  
CC useful for preventing and/or treating diseases related to schizophrenia.  
XX

XX Sequence 33780 BP; 9140 A; 6838 C; 7205 G; 10408 T; 189 other;

Query Match 1.2%; Score 20; DB 22; Length 33780;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 913 TTTTATTAACTCTCTTGTT 932  
|||||

DB 21958 TTTTATTAACTCTCTTGTT 21977

RESULT 24

ABK83575 standard; cDNA; 136284 BP.

XX ABK83575;

DT 14-AUG-2002 (first entry)

DE Human CDNA differentially expressed in granulocytic cells #146.

XX Human: ss: granulocytic cell; DNA chip: bacterial infection;  
XX viral infection; parasitic infection; protozoal infection;  
XX fungal infection; sterile inflammatory disease; psoriasis;  
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
XX cardiac reperfusion injury; renal reperfusion injury; ARDS;  
XX adult respiratory distress syndrome; inflammatory bowel disease;  
XX Crohn's disease; ulcerative colitis; periodontal disease;  
XX granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

PF 03-OCT-2001; 2001WO-US30821.

XX 03-OCT-2000; 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

DR WPI; 2002-435328/46.

XX  
PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity -

PS Claim 1: SEQ ID NO 146; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,

CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
CC periodontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 136284 BP; 41693 A; 28668 C; 28000 G; 37923 T; 0 other;

Query Match 1.2%; Score 20; DB 24; Length 136284;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 187 ATACAAAGACACTCTTATT 206  
|||||

DB 64587 ATACAAAGACACTCTTATT 64606

RESULT 25

AAH12882 standard; cDNA; 562 BP.

XX AAH12882;

DT 26-JUN-2001 (first entry)

DE Human CDNA clone (3'-primer) SEQ ID NO:9717.

XX Human: primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX Claim 3: SEQ ID 9717; 2537pp + CD ROM; English.  
PS The present invention describes primer sets for synthesizing 5602  
XX

CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

SQ Sequence 562 BP; 177 A; 113 C; 79 G; 182 T; 11 other;

Query Match 1.2%; Score 19; DB 22; Length 562;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 628 ATATTCTTACTAGCTTC 646  
 ||||||||||||||||  
 DB 381 ATATTCTTACTAGCTTC 399

RESULT 26  
 AAX20940  
 ID AAX20940 standard; DNA; 644 BP.

XX AAX20940;

DT 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of *Treponema pallidum*.

KW *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;  
 enzyme production; ds.

OS *Treponema pallidum*.

PN W09859034-A2.

XX 30-DEC-1998.

PF 23-JUN-1998; 98WO-US13041.

XX 24-JUN-1997; 97US-0050667.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Fraser CM;

DR WPI: 1999-081273/07.

PT New isolated *Treponema pallidum* nucleic acids - used to develop  
 PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of *T. pallidum* infections, particularly syphilis

PS Claim 1: Page 981-982; 1150pp; English.

CC AAX20500-21243 represent polynucleotide sequences from the genome of  
 CC *Treponema pallidum*. The sequences can be used for detection,  
 CC diagnosis, characterisation, prevention and therapy for *T. pallidum*  
 CC infections, particularly syphilis. They can also be used for detecting  
 CC diseases related to *Borrelia* infections in animals, and for the

CC production of biosynthetic products such as enzymes.  
 XX Sequence 644 BP; 206 A; 129 C; 128 G; 178 T; 3 other;

Query Match 1.2%; Score 19; DB 20; Length 644;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 835 AAAAGGGGAAAAAGGTGC 853  
 ||||||||||||||||  
 DB 345 AAAAGGGGAAAAAGGTGC 363

RESULT 27  
 AA196504/C  
 ID AA196504 standard; CDNA; 764 BP.

XX AA196504;

DT 13-NOV-2001 (first entry)

DE Human neuroblastoma expressed polynucleotide SEQ ID NO 2579.

KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.

OS Homo sapiens.

PN W0200166719-A1.

PD 13-SEP-2001.

PF 02-MAR-2001; 2001WO-JP01629.

XX 07-MAR-2000; 2000JP-0159195.

PA (CHTB-) CHIBA PREFECTURE.

PI (HISM) HISAMITSU PHARM CO LTD.

DR WPI: 2001-565584/63.

PT Nucleic acids originating in gene expressed in human neuroblastoma,  
 PT useful as probe or primer in diagnosing prognosis of human  
 PT neuroblastoma, malignancy and susceptibility indicator or tumour marker  
 PT for anti-cancer agents -

PS Claim 1: Page 1895; 2979pp; Japanese.

CC The invention relates to novel genes (AA193926-AA197963) expressed in  
 CC human neuroblastoma. The nucleic acids are applicable as a probe or  
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
 CC susceptibility indicators or tumour markers for anti-cancer agents. The  
 CC gene information for diagnosing prognosis is related to factors similar  
 CC to that for N-myc and TrkA genes.

SQ Sequence 764 BP; 217 A; 131 C; 158 G; 227 T; 31 other;

Query Match 1.2%; Score 19; DB 22; Length 764;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 704 TCCTGGGGAATTCTTTA 722  
 ||||||||||||||||  
 DB 422 TCCTGGGGAATTCTTTA 404

RESULT 28  
 AAS74111/C  
 ID AAS74111 standard; CDNA; 790 BP.

XX AAS74111;



DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #9915.  
 XX  
 DE Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR P-PSDB; ABG09924.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 1: SEQ ID NO 9915; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. A564197-A594564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 790 BP; 279 A; 132 C; 169 G; 210 T; 0 other;  
 XX  
 Query Match 1.2%; Score 19; DB 23; Length 790;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 OY 900 TCTCTTATATCTTTTAT 918  
 Db 728 TCTCTTATATCTTTTAT 710  
 XX  
 RESULT 29  
 AAF91438  
 ID AAF91438 standard; DNA; 1001 BP.  
 XX  
 AC AAF91438;  
 XX  
 DT 04-MAY-2001 (first entry)  
 XX

DE Haemophilus influenzae (Hird) Htrb gene upstream sequence, SEQ ID:64.  
 XX  
 KW Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;  
 KW genetically modified; protective antigen expression; LPS detoxification;  
 KW LPS; Lipid A; homologous recombination vector; immunisation;  
 KW immunoprotective; non-toxic; paediatric; ds.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO200109350-A2.  
 XX  
 PD 08-FEB-2001.  
 XX  
 PF 31-JUL-2000; 2000WO-EP07424.  
 XX  
 PR 03-AUG-1999; 99GB-0018319.  
 XX  
 PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Berthet FJ, Palemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;  
 PI Poolman J, Thiry G, Thonard J, Voet P;  
 XX  
 DR WPI: 2001-138654/14.  
 DR  
 XX  
 PT New isolated polynucleotide useful for outer membrane vesicle  
 PT preparation from Gram-negative bacterial strain for vaccination of  
 PT microbial infections -  
 XX  
 PS Claim 46: Page 93-94; 128bp; English.  
 XX  
 CC The invention relates to a genetically-engineered outer membrane vesicle  
 CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.  
 CC The blebs of the invention are improved with respect to their  
 CC immunogenicity and toxicity by the introduction of one or more genetic  
 CC changes to the chromosome of the bacterium from which the blebs are  
 CC derived. The changes made include the upregulation of protective antigen  
 CC expression, the downregulation of immunodominant non-protective antigen  
 CC expression, and genetic changes which result in detoxification of the  
 CC lipid A moiety of lipopolysaccharide (LPS). The invention also  
 CC encompasses modified Gram-negative bacterial strains from which the bleb  
 CC preparations are made, a vector suitable for performing recombination  
 CC events (for the generation of the modified bacterial strains),  
 CC bacterially-derived nucleic acid sequences used in such a vector, and an  
 CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole  
 CC cell vaccine suitable for paediatric use. The bleb preparation is useful  
 CC in the manufacture of a medicament for immunising a human host against a  
 CC disease caused by infection of one or more of the following: *Neisseria*  
 CC meningitidis, *Neisseria gonorrhoeae*, *Haemophilus influenzae*, *Moraxella*  
 CC catarrhalis, *Pseudomonas aeruginosa*, *Chlamydia trachomatis*, and *Chlamydia*  
 CC pneumoniae. The invention may also be used to provide immunisation against  
 CC the influenza virus. Bacterially derived nucleotide sequences of the  
 CC invention are used in the performance of homologous recombination events  
 CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either  
 CC increase or decrease expression of that gene. Immunoprotective and  
 CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines  
 CC are more immunogenic, less toxic and safer, and are particularly useful  
 CC for paediatric use. The present sequence represents a specifically  
 CC claimed *Haemophilus influenzae* nucleic acid sequence.  
 XX  
 SQ Sequence 1001 BP; 301 A; 225 C; 168 G; 307 T; 0 other;  
 XX  
 Query Match 1.2%; Score 19; DB 22; Length 1001;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 OY 150 TTCTCTGATTTACATTTA 168  
 Db 275 TTCTCTGATTTACATTTA 293  
 XX  
 RESULT 30  
 ABK37818  
 ID ABK37818 standard; DNA; 1001 BP.  
 XX



KW	Human; microglobulin transcriptional control factor 30;	
KW	recombinant production; malignant tumour; cancer; blood disease;	
KW	HIV infection; human immunodeficiency virus; immune disorder;	
KW	inflammatory condition; cytostatic; anti-HIV; antiinflammatory;	
KW	immunomodulator; ss.	
OS	Homo sapiens.	
XX		
FH	Key	location/Qualifiers
FT	CDS	99..908
FT		/*tag= a
FT		/product= "Human microglobulin transcriptional control
FT		factor 30"
XX		
PN	MO20018151-A1.	
XX		
PD	22-NOV-2001.	
XX		
PE	08-MAY-2001; 2001MO-CN00709.	
XX		
PR	09-MAY-2000; 2000CN-0115610.	
XX		
PA	(SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.	
XX		
PI	Mao Y, Xie Y;	
XX		
DR	WPI: 2002-034729/04.	
DR	P-PSDB; AAM52672.	
XX		
PT	Human Microglobulin Transcriptional Control Factor 30 and encoding	
PT	polynucleotide, used in diagnosis and treatment of malignant tumors,	
PT	hemopathy, human immunodeficiency virus infection, immunological	
PT	diseases and inflammation -	
XX		
PS	Claim 6; Page 30-31; 38pp; Chinese.	
XX		
CC	This sequence represents cDNA encoding human microglobulin	
CC	transcriptional control factor 30. The protein has a molecular weight of	
CC	30 kD. The invention relates to human microglobulin transcriptional	
CC	control factor 30 (AAM52672), nucleic acids encoding it (ABA02319), and	
CC	a method for the recombinant production of microglobulin transcriptional	
CC	control factor 30. The present invention additionally discloses an	
CC	antagonist of microglobulin transcriptional control factor 30 for	
CC	therapeutic use, and an antibody which specifically binds to	
CC	microglobulin transcriptional control factor 30. Microglobulin	
CC	transcriptional control factor 30, and nucleotides which encode it may	
CC	be used for treating a variety of diseases, such as malignant tumours,	
CC	blood diseases, HIV (human immunodeficiency virus) infection, immune	
CC	disorders and inflammatory conditions. The protein may also be used to	
CC	screen for modulators of its activity or for peptide fingerprinting	
CC	identification. The polynucleotide can be used as a primer for nucleic	
CC	acid amplification reactions or as a probe for hybridisation reactions,	
CC	or in producing gene chips or microarrays.	
XX		
SQ	Sequence 1604 BP; 572 A; 230 C; 269 G; 533 T; 0 other:	
	Query Match	1.2%; Score 19; DB 24; Length 1604;
	Best Local Similarity	100.0%; Pred. No. 53;
	Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	900 TCTCCTTATTTCTTTTAT 918	
DB	621 TCTCCTTATTTCTTTTAT 603	
RESULT 33		
ABLA1351		
XX	ABLA1351 standard; cDNA; 1689 BP.	
AC	ABLA1351;	
XX		
XX	22-MAY-2002. (first entry)	
XX		

DE	Human glycosyl-phosphatidyl inositol glycan F10.89 encoding cDNA.
XX	
XX	
XX	Human; glycosyl-phosphatidyl inositol glycan F10.89;
KW	embryonic development malformation; immunological dysfunction;
KW	hormonal metabolic disturbance; gene; ss.
XX	
OS	Homo sapiens.
XX	
XX	
XX	Key
FT	Location/Qualifiers
FT	CDS 945..1244
FT	/*tag= a
FT	/product= "glycosyl-phosphatidyl inositol glycan F10.89"
XX	
XX	CN1326960-A.
PN	
PD	19-DEC-2001.
XX	
PE	05-JUN-2000; 2000CN-0116326.
XX	
PR	05-JUN-2000; 2000CN-0116326.
XX	
PA	(BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX	
PI	Mao Y, Xie Y;
XX	
DR	WPI: 2002-206969/27.
DR	P-PSDB: ABB04400.
XX	
PT	New polypeptide-glycosyl-phosphatidyl inositol glycan F10.89 and
PT	polynucleotide encoding the polypeptide -
XX	
PS	Claim 6; Page 27-28 (Disclosure): 35pp; Chinese.
XX	
XX	The invention relates to human glycosyl-phosphatidyl inositol glycan
CC	F10.89, the polynucleotide encoding this polypeptide and DNA recombinant
CC	processes to produce the polypeptide. The present invention also
CC	discloses the method of applying the polypeptide in treating various
CC	diseases, such as embryonic development malformation, immunological
CC	dysfunction disease and hormonal metabolic disturbance disease. The
CC	present invention also discloses the antagonist for resisting the
CC	polypeptide and its treatment effect. The present invention also
CC	discloses the application of the polynucleotide for encoding
CC	glycosyl-phosphatidyl inositol glycan F10.89.
XX	
SQ	Sequence 1689 BP; 541 A; 345 C; 348 G; 455 T; 0 other;
XX	
Query Match	1.2%; Score 19; DB 24; Length 1689;
Best Local Similarity	100.0%; Pred. NO. 53;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	187 ATACAAGACACTCTTAT 205
Db	311 ATACAAGACACTCTTAT 329
XX	
RESULT 34	
AAH15049/c	
ID	AAH15049 standard; cDNA: 1833 BP.
XX	
AC	AAH15049;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA sequence SEQ ID NO:13039.
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
XD	07-FEB-2001.
XX	

PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and/or the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 13039; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification, where a primer set  
 CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 1833 BP; 595 A; 309 C; 313 G; 616 T; 0 other;  
 XX  
 Query Match 1.2%; Score 19; DB 22; Length 1833;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 628 ATATTCTTCTACTAGCTTC 646  
 ||||||||||||||||  
 DB 1453 ATATTCTTCTACTAGCTTC 1435  
 RESULT 35  
 ABL12876/C  
 ID ABL12876 standard; cDNA: 2611 BP.  
 XX  
 AC ABL12876;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33110.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX

PD 27-SEP-2001.  
 XX  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB68773.  
 XX  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 33110; 21pp + Sequence listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AAB57737-ABR72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2611 BP; 855 A; 546 C; 540 G; 670 T; 0 other;  
 XX  
 Query Match 1.2%; Score 19; DB 23; Length 2611;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 584 CCACGACGACAAAGCCTTC 602  
 ||||||||||||||||  
 DB 891 CCACGACGACAAAGCCTTC 873  
 RESULT 36  
 ABL04386/C  
 ID ABL04386 standard; cDNA: 3052 BP.  
 XX  
 AC ABL04386;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 7640.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB60283.  
 XX

XX	PT	New isolated nucleic acid detection reagent for detecting 1000 or more
XX	PT	genes from Drosophila and for elucidating cell signalling and cell-cell
XX	PT	interactions -
PS	XX	Claim 1; SEQ ID NO 7640; 21np + Sequence Listing; English.
CC	XX	The invention relates to an isolated nucleic acid detection reagent
CC	CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	CC	useful in developmental biology and in elucidating cell signalling and
CC	CC	cell-cell interactions in higher eukaryotes for the development of
CC	CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	CC	(ABB57737-ABB72072).
CC	CC	The sequence data for this patent did not form part of the printed
CC	CC	specification, but was obtained in electronic format directly from WIPO
CC	CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	XX	Sequence 3052 BP; 792 A; 651 C; 705 G; 904 T; 0 other;
QY	Db	299 AAGCTAAGTAATATTTTA 317
		591 AAGCTAAGTAATATTTTA 573
RESULT 37		
ABK13001/C		
ID	ABK13001	standard; cDNA; 3752 BP.
XX	AC	
XX	ABK13001;	
DT	23-APR-2002	(first entry)
DE	DNA encoding SCAN zinc-finger protein subfamily protein 57.	
XX	XX	
XX	SCAN zinc-finger protein subfamily protein 57; cytosolic;	
KW	neuroprotectant; neurotropic; cerebroprotective; tumour;	
KW	nervous system disorder; neurodegenerative disease; neural disease;	
KW	developmental disturbance; psychogenic disease; gene; ss.	
XX	XX	
OS	Unidentified.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	131..1687
FT		/*tag= a
FT		/product= "SCAN zinc-finger protein subfamily protein 57"
XX	PN	WO200194530-A2.
XX	XX	
XX	PD	13-DEC-2001.
XX	PE	14-MAY-2001; 2001WO-CN00774.
XX	XX	
PR	16-MAY-2000; 2000CN-0115706.	
PA	(SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.	
PI	Mao Y, Xie Y;	
XX	XX	
DR	WI: 2002-083223/11.	
DR	P-PSDB; AAU11020.	
XX	XX	
XX	SCAN zinc-finger subfamily protein 57 and encoding polynucleotide, for	
XX	diagnosis and treatment of tumours of the nervous system,	
XX	neurodegenerative disease, developmental disturbances, psychogenic	
XX	disease and other neural diseases -	
XX	XX	
XX	Claim 6; Page 30-32; 35pp; Chinese	

XX	The invention relates to an isolated polypeptide (I) of SCAN zinc-finger
CC	protein subfamily protein 57 and the polynucleotide (I) encoding (II).
CC	(I) and (II) are used in diagnosis and treatment of tumours of the
CC	nervous system, neurodegenerative disease, developmental disturbances
CC	of the nervous system, psychogenic disease and other neural diseases.
CC	The present sequence represents the coding sequence of SCAN zinc-
CC	finger protein subfamily protein 57.
XX	
SQ	Sequence 3752 BP; 1205 A; 704 C; 754 G; 1089 T; 0 other;
Query Match	1.2%; Score 19; DB 24; Length 3752;
Best Local Similarity	100.0%; Pred. No. 52;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	628 ATATTCTTTACTAGCTTC 646 
Db	3376 ATATTCTTTACTAGCTTC 3358
RESULT 38	
ID	ABL24540/C
XX	ABL24540 standard; DNA; 4007 BP.
AC	ABL24540;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 25093.
XX	
KM	Drosophila: developmental biology; cell signalling; insecticide;
KW	Pharmaceutical; gene; ds.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO20011042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE ) PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
XX	interactions -
PS	Claim 1; SEQ ID NO 25093; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC	sequences (AB101840-AB16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 4007 BP; 1201 A; 929 C; 887 G; 990 T; 0 other;
Query Match	1.2%; Score 19; DB 23; Length 4007;
Best Local Similarity	100.0%; Pred. No. 52;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

QY 584 CCACGAGCAAGCGTTTC 602  
 |||||||  
 DB 3749 CCACGAGCAAGCGTTTC 3731

RESULT 39  
 ABL11410/c  
 ID ABL11410 standard; cDNA; 5586 BP.  
 XX  
 AC ABL11410;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 28712.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR P-PSDB; ABB67307.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 28712; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SO Sequence 5586 BP; 1581 A; 1196 C; 1164 G; 1645 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 5586;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 CCTTCATCAGTTTGGC 258  
 |||||||  
 DB 913 CCTTCATCAGTTTGGC 895

RESULT 40  
 ABL03008/c  
 ID ABL03008 standard; cDNA; 8034 BP.  
 XX  
 AC ABL03008;  
 XX  
 DT 26-MAR-2002 (first entry)

XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 3506.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR P-PSDB; ABB58905.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 3506; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SO Sequence 8034 BP; 2137 A; 1815 C; 1882 G; 2200 T; 0 other;

Query Match 1.2%; Score 19; DB 23; Length 8034;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 299 AAGCTAGTAATATTTTA 317  
 |||||||  
 DB 6827 AAGCTAGTAATATTTTA 6809

Search completed: January 21, 2003, 21:03:30  
 Job time : 2525 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 21:03:19 : Search time 59 seconds  
(without alignments)  
8508.981 Million cell updates/sec

Title: US-09-869-433-1

Perfect score: 1637  
Sequence: 1 gaataaaactatcagaa.....gttgatccctactcgtcttt 1637

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCITUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	1.3	8220	2	US-08-568-459A-11
2	21	1.3	8220	2	US-08-487-826B-11
3	21	1.3	8220	4	US-09-210-288-11
4	21	1.3	19124	2	US-08-487-826B-13
5	18	1.1	3980	1	US-08-233-008A-1
6	18	1.1	3980	1	US-08-233-008A-5
7	18	1.1	6321	4	US-09-221-017B-311
8	18	1.1	6519	1	US-08-233-008A-7
9	18	1.1	14736	4	US-08-961-527-171
10	17	1.0	378	4	US-09-134-001C-2567
11	17	1.0	383	4	US-09-370-838-285
12	17	1.0	882	4	US-09-227-357-133
13	17	1.0	939	1	US-08-592-411-14
14	17	1.0	939	1	US-08-591-401-15
15	17	1.0	942	1	US-08-592-411-16
16	17	1.0	942	1	US-08-591-501-14
17	17	1.0	1316	2	US-08-871-924A-1
18	17	1.0	1316	4	US-09-361-900-1
19	17	1.0	1374	1	US-08-093-372-3
20	17	1.0	1380	2	US-08-748-947A-1
21	17	1.0	1614	4	US-09-110-959A-103
22	17	1.0	2561	4	US-09-270-542-101
23	17	1.0	2561	4	US-09-270-542-119
24	17	1.0	2880	1	US-08-158-189-1
25	17	1.0	3099	4	US-09-360-186-2
26	17	1.0	3810	2	US-08-475-844-8
27	17	1.0	3810	5	PCR-US95-08429-8





```
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210.288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 8220 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-11

Query Match      1.3%; Score 21; DB 4; Length 8220;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAACCTTTGGAAAT 84
|||||
Db 5046 AAGAAAACCTTTGGAAAT 5066

RESULT 4
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Shim, Kim I.
APPLICANT: Chlinitz, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487.826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
```

```
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match      1.3%; Score 21; DB 2; Length 19124;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 AAGAAAACCTTTGGAAAT 84
|||||
Db 12366 AAGAAAACCTTTGGAAAT 12386

RESULT 5
US-08-233-008A-1
Sequence 1, Application US/08233008A
Patent No. 5578480
GENERAL INFORMATION:
APPLICANT: Khandke, Kiran M.
TITLE OF INVENTION: Methods For The Isolation And
TITLE OF INVENTION: Purification Of The Recombinantly Expressed chondroitinase
TITLE OF INVENTION: I and II Enzymes From P. vulgaris
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07470-8426
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.008A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 31,885-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3980 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 119..3181
US-08-233-008A-1

Query Match      1.1%; Score 18; DB 1; Length 3980;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 191 AAAGACACTCTTATGT 208
|||||
Db 2965 AAAGACACTCTTATGT 2982
```

```
RESULT 6
US-08-233-008A-5
; Sequence 5, Application US/08233008A
; Patent No. 5578480
; GENERAL INFORMATION:
; APPLICANT: Khandke, Kiran M.
; TITLE OF INVENTION: Methods For The Isolation And
; TITLE OF INVENTION: Purification Of The Recombinantly Expressed chondroitinase
; TITLE OF INVENTION: I and II Enzymes From P. vulgaris
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,008A
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 31,885-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3244
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3980 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 188..3181
US-08-233-008A-5

Query Match 1.1%; Score 18; DB 1; Length 3980;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 AAAAGACACTTATTGT 208
Db 2965 AAAAGACACTTATTGT 2982

RESULT 7
US-09-221-017B-311
; Sequence 311, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/A098/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 311:
SEQUENCE CHARACTERISTICS:
LENGTH: 6321 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...6321
US-09-221-017B-311

Query Match 1.1%; Score 18; DB 4; Length 6321;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1163 TTCTTCGCTCTTGTAT 1180
Db 6152 TTCTTCGCTCTTGTAT 6169

RESULT 8
US-08-233-008A-7
; Sequence 7, Application US/08233008A
; Patent No. 5578480
; GENERAL INFORMATION:
; APPLICANT: Khandke, Kiran M.
; TITLE OF INVENTION: Methods For The Isolation And
; TITLE OF INVENTION: Purification Of The Recombinantly Expressed chondroitinase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,008A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gordon, Alan M.
REGISTRATION NUMBER: 30,637
REFERENCE/DOCKET NUMBER: 31,885-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3244
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6519 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 3238..6276
US-08-233-008A-7

Query Match          1.1%; Score 18; DB 1: Length 6519;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 AAAAGACACTTATTGT 208
    |||||
DB 2965 AAAAGACACTTATTGT 2982

RESULT 9
US-08-961-527-171
Sequence 171, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 171:
SEQUENCE CHARACTERISTICS:
LENGTH: 14736 base pairs
TYPE: nucleic acid
```

```
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-171

Query Match          1.1%; Score 18; DB 4: Length 14736;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 ATCTACAGAAATTTGCTG 429
    |||||
DB 5312 ATCTACAGAAATTTGCTG 5329

RESULT 10
US-09-134-001C-2567
Sequence 2567, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2567
LENGTH: 378
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2567

Query Match          1.0%; Score 17; DB 4: Length 378;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1549 TTAAGACACAGAGAGTG 1565
    |||||
DB 29 TTAAGACACAGAGAGTG 45

RESULT 11
US-09-370-838-285/c
Sequence 285, Application US/09370838
Patent No. 6444425
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohamath, Roadoh
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370,838
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/285,323
EARLIER FILING DATE: 1999-04-02
NUMBER OF SEQ ID NOS: 289
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 285
LENGTH: 383
TYPE: DNA
ORGANISM: Homo sapiens
US-09-370-838-285

Query Match          1.0%; Score 17; DB 4: Length 383;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 571 AAATTACAAAAATCCAC 587
```

Db 367 AATTACAAATTCAC 351

```

|||||
RESULT 12
US-09-227-357-133/C
; Sequence 133, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: p2010p1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785

```

```

; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO: 133
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (881)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-133

```

Query Match 1.0%; Score 17; DB 4; Length 882;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 ACCTAAGCAAGTCTG 134  
Db 875 ACCTAAGCAAGTCTG 859

```

RESULT 13
US-08-592-411-14/C
; Sequence 14, Application US/08592411
; Patent No. 5726032
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for the Efficient Production of
; TITLE OF INVENTION: 7-ADCA via 2-(Carboxymethylthio)sceyl-7-ADCA and
; NUMBER OF SEQUENCES: 17
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,411
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 939 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: No. 5726032ardia lactandurens
; INDIVIDUAL ISOLATE: ATCC 27382
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..939
; OTHER INFORMATION: /gene="cefE"
US-08-592-411-14

```

Query Match 1.0%; Score 17; DB 1; Length 939;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1366 TTGATGATGTCGCC 1382  
Db 914 TTGATGATGTCGCC 898

```
RESULT 14
US-08-591-501-15/C
; Sequence 15, Application US/08591501
; Patent No. 5795733
; GENERAL INFORMATION:
; APPLICANT: BOVENBERG, ROELOF ARY LANS
; APPLICANT: KOEKMAN, BERTUS PIETER
; APPLICANT: HOEKEMA, ANDREAS
; APPLICANT: VAN DER LAAN, JAN METSKE
; APPLICANT: VERWEIJ, JAN
; APPLICANT: DE VROOM, ERIK
; TITLE OF INVENTION: PROCESS FOR THE EFFICIENT PRODUCTION OF
; TITLE OF INVENTION: 7-ADCA VIA 3-(CARBOXYETHYLTHIO) PROPIONYL-7-ADCA
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE, NW
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,501
; FILING DATE: 13-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30, 988
; REFERENCE/DOCKET NUMBER: 24615-20065.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 822-0168
; TELEX: 90-4030 MRSNFOERSM
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 939 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-591-501-15
Query Match 1.0%; Score 17; DB 1; Length 939;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 TTGATGTAGTGGCCGC 1382
DB 914 TTGATGTAGTGGCCGC 898

RESULT 15
US-08-592-411-16/C
; Sequence 16, Application US/08592411
; Patent No. 5726032
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for the efficient production of
; TITLE OF INVENTION: 7-ADCA via 2-(Carboxyethylthio)acetyl-7-ADCA and
; TITLE OF INVENTION: 3-(Carboxymethylthio)propionyl-7-ADCA
; NUMBER OF SEQUENCES: 17
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,411
```

```
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 942 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: No. 5726032ardia lactamdurans
; INDIVIDUAL ISOLATE: LC 411
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..942
; OTHER INFORMATION: /gene="cefE"
; OTHER INFORMATION: /citation=({1})
; PUBLICATION INFORMATION:
; AUTHORS: Coque, J J
; AUTHORS: Martin, J F
; AUTHORS: Liras, P
; TITLE: Characterization and expression in
; TITLE: Streptomyces lividans of cefD and cefE genes from
; TITLE: No. 5726032ardia lactamdurans: the organization of the
; TITLE: cephamycin gene cluster differs from that in
; TITLE: Streptomyces clavuligerus
; JOURNAL: Mol. Gen. Genet.
; VOLUME: 236
; PAGES: 453-458
; DATE: 1993
;
US-08-592-411-16
Query Match 1.0%; Score 17; DB 1; Length 942;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1366 TTGATGTAGTGGCCGC 1382
DB 917 TTGATGTAGTGGCCGC 901

RESULT 16
US-08-591-501-14/C
; Sequence 14, Application US/08591501
; Patent No. 5795733
; GENERAL INFORMATION:
; APPLICANT: BOVENBERG, ROELOF ARY LANS
; APPLICANT: KOEKMAN, BERTUS PIETER
; APPLICANT: HOEKEMA, ANDREAS
; APPLICANT: VAN DER LAAN, JAN METSKE
; APPLICANT: VERWEIJ, JAN
; APPLICANT: DE VROOM, ERIK
; TITLE OF INVENTION: PROCESS FOR THE EFFICIENT PRODUCTION OF
; TITLE OF INVENTION: 7-ADCA VIA 3-(CARBOXYETHYLTHIO) PROPIONYL-7-ADCA
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE, NW
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,501
; FILING DATE: 13-MAY-1996
; CLASSIFICATION: 435
```

ATTORNEY/AGENT INFORMATION:  
NAME: ADLER, REID G.  
REGISTRATION NUMBER: 30,988  
REFERENCE/DOCKET NUMBER: 24615-20065.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 822-0168  
TELEX: 90-4030 MRSNFOERSWSH  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 942 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PUBLICATION INFORMATION:  
AUTHORS: COQUE et al.,  
JOURNAL: Mol. Gen. Genet.  
VOLUME: 236  
PAGES: 453-458  
DATE: 1993  
RELEVANT RESIDUES IN SEQ ID NO: 14: FROM 1 TO 942  
US-08-591-501-14

Query Match 1.0%; Score 17; DB 1; Length 942;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1366 TTGATGTAGTTGCCGCC 1382  
DB 917 TTGATGTAGTTGCCGCC 901

RESULT 17  
US-08-871-924A-1/c  
Sequence 1, Application US/08871924A  
Patent No. 5932469  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jeffrey D  
TITLE OF INVENTION: No. 5932469e1 Antimicrobial Polypeptides and  
TITLE OF INVENTION: Methods of Use  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney St  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASC II/DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,924A  
FILING DATE: 06/10/97  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 920308.90649  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1316 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: -35\_signal

LOCATION: 738..742  
FEATURE:  
NAME/KEY: -10\_signal  
LOCATION: 757..763  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 784..791  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 796..987  
US-08-871-924A-1

Query Match 1.0%; Score 17; DB 2; Length 1316;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 863 AATGAATATGAAGAATA 879  
DB 260 AATGAATATGAAGAATA 244

RESULT 18  
US-09-361-900-1/c  
Sequence 1, Application US/09361900  
Patent No. 6391285  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jeffrey D  
TITLE OF INVENTION: No. 6391285e1 Antimicrobial Polypeptides and  
TITLE OF INVENTION: Methods of Use  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney St  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASC II/DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/361,900  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/871,924  
FILING DATE: 06/10/97  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 920308.90649  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1316 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: -35\_signal  
LOCATION: 738..742  
FEATURE:  
NAME/KEY: -10\_signal  
LOCATION: 757..763  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 784..791  
FEATURE:

NAME/KEY: CDS  
LOCATION: 796..987  
US-09-361-900-1

Query Match 1.0%; Score 17; DB 4; Length 1316;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 863 AATGATATGAAGAATA 879  
|||||  
DB 260 AATGATATGAAGAATA 244

RESULT 19  
US-08-093-372-3/c  
Sequence 3, Application US/08093372  
Patent No. 5530187  
GENERAL INFORMATION:  
APPLICANT: Lamb, Christopher J.  
APPLICANT: Zhu, Qun  
APPLICANT: Maher, Eileen A.  
APPLICANT: Dixon, Richard A.  
TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING MULTIPLE  
TITLE OF INVENTION: DISEASE RESISTANCE GENES  
NUMBER OF SEQUENCES: 4  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/093.372  
FILING DATE: 16-JUL-1993  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1374 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 52..1158  
OTHER INFORMATION: /product= "ALFALFA BETA-GLUCANASE"  
US-08-093-372-3

Query Match 1.0%; Score 17; DB 1; Length 1374;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 861 AAATGATATGAAGA 877  
|||||  
DB 28 AAATGATATGAAGA 12

RESULT 20  
US-08-748-947A-1/c  
Sequence 1, Application US/08748947A

Patent No. 5854031  
GENERAL INFORMATION:  
APPLICANT: Jigami et al., Yoshifumi  
TITLE OF INVENTION: MANNOSE-1-PHOSPHATE TRANSFERASE GENE  
TITLE OF INVENTION: FROM YEAST, AND A PROCESS FOR PRODUCING  
TITLE OF INVENTION: MANNOSE-1-PHOSPHATE-CONTAINING ACIDIC SUGAR CHAINS AND  
TITLE OF INVENTION: PHOSPHATE-CONTAINING ACIDIC SUGAR CHAINS BY USE OF THE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 601 Thirteenth Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/748.947A  
FILING DATE: 14-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 299509/1995  
FILING DATE: 17-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ellison, Eldora L.  
REGISTRATION NUMBER: 39,967  
REFERENCE/DOCKET NUMBER: 08206/006001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/783-5070  
TELEFAX: 202/783-2331  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1380 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1339  
US-08-748-947A-1

Query Match 1.0%; Score 17; DB 2; Length 1380;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 560 ATTGCTAATGAATTA 576  
|||||  
DB 648 ATTGCTAATGAATTA 632

RESULT 21  
US-09-110-959A-3/c  
Sequence 3, Application US/09110959A  
Patent No. 6268197  
GENERAL INFORMATION:  
APPLICANT: Schuelein, Martin  
APPLICANT: Outtrup, Helle  
APPLICANT: Jorgensen, Per Lina  
APPLICANT: Bjornvad, Mads Eskelund  
TITLE OF INVENTION: Alkaline xyloglucanase  
FILE REFERENCE: 5206.200-US  
CURRENT APPLICATION NUMBER: US/09/110.959A  
CURRENT FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 0822/97  
PRIOR FILING DATE: 1997-07-07  
PRIOR APPLICATION NUMBER: 1213/97  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/054,039

;; PRIOR FILING DATE: 1997-07-28  
;; PRIOR APPLICATION NUMBER: 60/063,694  
;; PRIOR FILING DATE: 1997-10-28  
;; NUMBER OF SEQ ID NOS: 14  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 3  
;; LENGTH: 1614  
;; TYPE: DNA  
;; ORGANISM: Bacillus agaradhaerens NCIMB 40482  
US-09-110-959A-3

Query Match 1.0%; Score 17; DB 4; Length 1614;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 913 TTTTAACTCTCTTG 929  
|||  
DB 144 TTTTAACTCTCTTG 128

RESULT 22  
US-09-270-542-101/c  
; Sequence 101, Application US/09270542  
; Patent No. 6322976  
; GENERAL INFORMATION:  
; APPLICANT: Altman, Timothy  
; APPLICANT: Scott, James  
; APPLICANT: Stanton, Lawrence  
; TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and  
; TITLE OF INVENTION: Therapy  
; FILE REFERENCE: 4198/78179  
; CURRENT APPLICATION NUMBER: US/09/270,542  
; CURRENT FILING DATE: 1999-03-17  
; EARLIER APPLICATION NUMBER: 09/221,222  
; EARLIER FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 207  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 101  
; LENGTH: 2561  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-09-270-542-101

Query Match 1.0%; Score 17; DB 4; Length 2561;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 CCTTTTAATTTCTTT 368  
|||  
DB 1444 CCTTTTAATTTCTTT 1428

RESULT 23  
US-09-270-542-119/c  
; Sequence 119, Application US/09270542  
; Patent No. 6322976  
; GENERAL INFORMATION:  
; APPLICANT: Altman, Timothy  
; APPLICANT: Scott, James  
; APPLICANT: Stanton, Lawrence  
; TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and  
; TITLE OF INVENTION: Therapy  
; FILE REFERENCE: 4198/78179  
; CURRENT APPLICATION NUMBER: US/09/270,542  
; CURRENT FILING DATE: 1999-03-17  
; EARLIER APPLICATION NUMBER: 09/221,222  
; EARLIER FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 207  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 119  
; LENGTH: 2561  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus

US-09-270-542-119

Query Match 1.0%; Score 17; DB 4; Length 2561;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 352 CCTTTTAATTTCTTT 368  
|||  
DB 1444 CCTTTTAATTTCTTT 1428

RESULT 24  
US-08-158-189-1  
; Sequence 1, Application US/08158189  
; Patent No. 5641497  
; GENERAL INFORMATION:  
; APPLICANT: Bevins, Charles L.  
; APPLICANT: Jones, Douglas E.  
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,  
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: U.S.A.  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/158,189  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/888,232  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Johnson, Philip S.  
; REGISTRATION NUMBER: 27,200  
; REFERENCE/DOCKET NUMBER: CH-0219  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2880 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1399..1572  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2552..2663  
; FEATURE:  
; NAME/KEY: TATA\_signal  
; LOCATION: 1328..1334  
; FEATURE:  
; NAME/KEY: CAAT\_signal  
; LOCATION: 1267..1271  
; FEATURE:  
; NAME/KEY: misc\_signal  
; LOCATION: 1569..1576  
; FEATURE:  
; NAME/KEY: misc\_signal  
; LOCATION: 2539..2549  
; FEATURE:



```

: NAME/KEY: polyA-signal
: LOCATION: 2770..2775
US-08-158-189-1
:
Query Match 1.0%; Score 17; DB 1; Length 2880;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 CCTTTTAATTTCTTT 368
Db 2804 CCTTTTAATTTCTTT 2820

RESULT 25
US-09-360-186-2/c
: Sequence 2, Application US/09360186
: Patent No. 6262343
: GENERAL INFORMATION:
: APPLICANT: Staskawicz, et al.
: TITLE OF INVENTION: Bst Resistance Gene
: FILE REFERENCE: 50687
: CURRENT APPLICATION NUMBER: US/09/360,186
: CURRENT FILING DATE: 1999-07-23
: EARLIER APPLICATION NUMBER: 60/093,957
: EARLIER FILING DATE: 1998-07-23
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 3099
: TYPE: DNA
: ORGANISM: Capsicum annuum
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (93)..(2810)
US-09-360-186-2

Query Match 1.0%; Score 17; DB 4; Length 3099;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 TTCTTAATGTTCTCTG 157
Db 2813 TTCTTAATGTTCTCTG 2797

RESULT 26
US-08-475-844-8
: Sequence 8, Application US/08475844
: Patent No. 5972643
: GENERAL INFORMATION:
: APPLICANT: Lobanenko, Victor V.
: APPLICANT: Nelman, Paul E.
: APPLICANT: Klenova, Elena M.
: APPLICANT: Goodwin, Graham H.
: APPLICANT: Filippova, Galina N.
: APPLICANT: Collins, Steven J.
: TITLE OF INVENTION: CTCF
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Townsend and Townsend Kourlie and Crew
: STREET: One Market Plaza
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,844
: FILING DATE: 07-JUN-1995
```

```

: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/261,680
: FILING DATE: 17-JUN-1994
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Parmelee, Steven W.
: REGISTRATION NUMBER: 31,990
: REFERENCE/DOCKET NUMBER: 14538A-11-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-467-9600
: TELEFAX: 415-543-5043
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3810 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: CLONE: human CTCF cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 292..2475
: FEATURE:
: NAME/KEY: exon
: LOCATION: 281..1074
: OTHER INFORMATION: /label= exon2
: FEATURE:
: NAME/KEY: exon
: LOCATION: 1075..1245
: OTHER INFORMATION: /label= exon3
: FEATURE:
: NAME/KEY: exon
: LOCATION: 1246..1379
: OTHER INFORMATION: /label= exon4
: FEATURE:
: NAME/KEY: exon
: LOCATION: 1380..1499
: OTHER INFORMATION: /label= exon5
: FEATURE:
: NAME/KEY: exon
: LOCATION: 1500..1649
: OTHER INFORMATION: /label= exon6
: FEATURE:
: NAME/KEY: exon
: LOCATION: 1650..1810
: OTHER INFORMATION: /label= exon7
: NAME/KEY: exon
: LOCATION: 1810..1992
: OTHER INFORMATION: /label= exon8
US-08-475-844-8

Query Match 1.0%; Score 17; DB 2; Length 3810;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 ATTTCTTGCCCTGTT 376
Db 3311 ATTTCTTGCCCTGTT 3327

RESULT 27
PCT-US95-08429-8
: Sequence 8, Application PC/TUS9508429
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: CTCF
: NUMBER OF SEQUENCES: 21
: COMPUTER READABLE FORM:
```

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08429  
FILING DATE: 15-JUN-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/261,680  
FILING DATE: 17-JUN-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 14538A-11-1PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-467-9600  
TELEFAX: 415-543-5043  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3810 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: human CTCF CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 292..2475  
FEATURE:  
NAME/KEY: exon  
LOCATION: 281..1074  
OTHER INFORMATION: /label= exon2  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1075..1245  
OTHER INFORMATION: /label= exon3  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1246..1379  
OTHER INFORMATION: /label= exon4  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1380..1499  
OTHER INFORMATION: /label= exon5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1500..1649  
OTHER INFORMATION: /label= exon6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1650..1810  
OTHER INFORMATION: /label= exon7  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1810..1992  
OTHER INFORMATION: /label= exon8  
PCT-US95-08429-8

Query Match 1.0%; Score 17; DB 5; Length 3810;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 ATTTCTTTGCCCTGTT 376  
|||||

Db 3311 ATTTCTTTGCCCTGTT 3327

RESULT 28

US-09-221-235-1/c  
; Sequence 1, Application US/09221235  
; Patent No. 6043040  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan  
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR  
; FILE REFERENCE: MNI-050  
; CURRENT FILING DATE: US/09/221,235  
; EARLIER FILING DATE: 1998-12-28  
; EARLIER APPLICATION NUMBER: 09/163,115  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4137  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid  
; NAME/KEY: CDS  
; LOCATION: (297)..(1202)  
US-09-221-235-1

Query Match 1.0%; Score 17; DB 3; Length 4137;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAAATAAAGTATTTC A 40  
|||||

Db 4005 AAAATAAAGTATTTC A 3989

RESULT 29

US-09-221-928-1/c  
; Sequence 1, Application US/09221928  
; Patent No. 6121030  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan  
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR  
; FILE REFERENCE: MNI-050  
; CURRENT APPLICATION NUMBER: US/09/221,928  
; CURRENT FILING DATE: 1998-12-28  
; EARLIER FILING DATE: 09/163,115  
; EARLIER APPLICATION NUMBER: 09/163,115  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4137  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid  
; NAME/KEY: CDS  
; LOCATION: (297)..(1202)  
US-09-221-928-1

Query Match 1.0%; Score 17; DB 3; Length 4137;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAAATAAAGTATTTC A 40  
|||||

Db 4005 AAAATAAAGTATTTC A 3989

RESULT 30

US-09-221-527-1/c  
; Sequence 1, Application US/09221527  
; Patent No. 6146832  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan

```
; TITLE OF INVENTION: NOVEL CSAK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,527
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4137
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
; NAME/KEY: CDS
; LOCATION: (297)..(1202)
US-09-221-527-1

Query Match          1.0%; Score 17; DB 3; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAAATAAAGTATTCA 40
DB 4005 AAAATAAAGTATTCA 3989

RESULT 31
US-09-221-236-1/c
; Sequence 1, Application US/09221236
; Patent No. 6146841
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,236
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4137
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
; NAME/KEY: CDS
; LOCATION: (297)..(1202)
US-09-221-236-1

Query Match          1.0%; Score 17; DB 3; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAAATAAAGTATTCA 40
DB 4005 AAAATAAAGTATTCA 3989

RESULT 32
US-09-221-416-1/c
; Sequence 1, Application US/09221416
; Patent No. 6153417
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,416
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4137
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
; NAME/KEY: CDS
; LOCATION: (297)..(1202)
US-09-221-416-1

Query Match          1.0%; Score 17; DB 3; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAAATAAAGTATTCA 40
DB 4005 AAAATAAAGTATTCA 3989

RESULT 33
US-09-221-245-1/c
; Sequence 1, Application US/09221245
; Patent No. 6180358
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,245
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: US 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4137
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
; NAME/KEY: CDS
; LOCATION: (297)..(1202)
US-09-221-245-1

Query Match          1.0%; Score 17; DB 4; Length 4137;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAAATAAAGTATTCA 40
DB 4005 AAAATAAAGTATTCA 3989

RESULT 34
US-09-163-115-1/c
; Sequence 1, Application US/09163115A
; Patent No. 6183962
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/163,115A
; CURRENT FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4137
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
FEATURE:
: OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (297)..(1202)
US-09-163-115-1
```

```
Query Match
Best Local Similarity 100.0%; Score 17; DB 4; Length 4137;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 24 AAAATAAAGTATTCA 40
Db 4005 AAAATAAAGTATTCA 3989
```

```
RESULT 35
US-09-221-528-1/c
: Sequence 1, Application US/09221528
: Patent No. 6190874
: GENERAL INFORMATION:
: APPLICANT: Acton, Susan
: TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
: FILE REFERENCE: MNI-050
: CURRENT APPLICATION NUMBER: US/09/221,528
: CURRENT FILING DATE: 1998-12-28
: EARLIER APPLICATION NUMBER: 09/163,115
: EARLIER FILING DATE: 1998-09-29
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4137
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
: NAME/KEY: CDS
: LOCATION: (297)..(1202)
US-09-221-528-1
```

```
Query Match
Best Local Similarity 100.0%; Score 17; DB 4; Length 4137;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 24 AAAATAAAGTATTCA 40
Db 4005 AAAATAAAGTATTCA 3989
```

```
RESULT 36
US-09-593-553-1/c
: Sequence 1, Application US/09593553
: Patent No. 6200770
: GENERAL INFORMATION:
: APPLICANT: Acton, Susan
: TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
: FILE REFERENCE: MNI-050
: CURRENT APPLICATION NUMBER: US/09/593,553
: CURRENT FILING DATE: 2000-06-14
: PRIOR APPLICATION NUMBER: 09/163,115
: PRIOR FILING DATE: 1998-09-28
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4137
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
: NAME/KEY: CDS
: LOCATION: (297)..(1202)
US-09-593-553-1
```

```
Query Match
Best Local Similarity 100.0%; Score 17; DB 4; Length 4137;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 24 AAAATAAAGTATTCA 40
Db 4005 AAAATAAAGTATTCA 3989
```

```
RESULT 37
US-09-221-237-1/c
: Sequence 1, Application US/09221237
: Patent No. 6214597
: GENERAL INFORMATION:
: APPLICANT: Acton, Susan
: TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
: FILE REFERENCE: MNI-050
: CURRENT APPLICATION NUMBER: US/09/221,237
: CURRENT FILING DATE: 1998-12-28
: EARLIER APPLICATION NUMBER: 09/163,115
: EARLIER FILING DATE: 1998-09-29
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4137
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: 'n' at positions 2872, 3597 and 3682 may be any nucleic acid
: NAME/KEY: CDS
: LOCATION: (297)..(1202)
US-09-221-237-1
```

```
Query Match
Best Local Similarity 100.0%; Score 17; DB 4; Length 4137;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 24 AAAATAAAGTATTCA 40
Db 4005 AAAATAAAGTATTCA 3989
```

```
RESULT 38
US-09-049-671-2/c
: Sequence 2, Application US/09049671
: Patent No. 5928674
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl J.
: APPLICANT: Baughn, Mariah
: TITLE OF INVENTION: NEKI-RELATED PROTEIN KINASE
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/049,671
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
```

```

: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0492 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4170 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: UTRSNOR01
: CLONE: 3069734
: US-09-049-671-2.

Query Match          1.0%; Score 17; DB 2; Length 4170;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAAATAAAGTATTTC A 40
Db 4046 AAAATAAAGTATTTC A 4030

RESULT 39
US-09-295-068-2/c
: Sequence 2, Application US/09295068
: Patent No. 6030801
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl J.
: APPLICANT: Baughn, Mariah
: TITLE OF INVENTION: NEKI-RELATED PROTEIN KINASE
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESS: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/295,068
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/049,671
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0492 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4170 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: UTRSNOR01

: CLONE: 3069734
: US-09-295-068-2

Query Match          1.0%; Score 17; DB 3; Length 4170;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 AAAATAAAGTATTTC A 40
Db 4046 AAAATAAAGTATTTC A 4030

RESULT 40
US-08-853-310-3/c
: Sequence 3, Application US/08853310
: Patent No. 5948640
: GENERAL INFORMATION:
: APPLICANT: Randazzo, Filippo
: TITLE OF INVENTION: Mammalian Additional Sex Combs (Asx) Acts as a Tumor Suppre
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Chilton Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 94608
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/853,310
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Gutb, Joseph H.
: REGISTRATION NUMBER: 31,261
: REFERENCE/DOCKET NUMBER: 1228,003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 923-3888
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5362 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-853-310-3

Query Match          1.0%; Score 17; DB 2; Length 5362;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1466 TGCAGTATCTTCTTT 1482
Db 1207 TGCAGTATCTTCTTT 1191

Search completed: January 21, 2003, 21:46:50
Job time : 187 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 17:12:48 : Search time 2947 Seconds

(without alignments) 16166.031 Million cell) updates/sec

Title: US-09-869-433-1

Sequence: 1 gaataaaactatcagaa.....gtgtacctactcgtcttt 1637

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ot:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_hum:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pin:*
35: em_hlg_rod:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1633.8	99.8	13389	1	AE002202	AE002202 Chlamydia
2	1632.2	99.7	11421	1	AE001619	AE001619 Chlamydia
3	1632.2	99.7	300650	1	AP002546	AP002546 Chlamydia
4	883.4	54.0	14168	1	AE001281	AE001281 Chlamydia
5	865.2	52.9	1587	1	TAJ10586	TAJ10586 Chlamydia
6	862.6	52.7	23773	1	AE002301	AE002301 Chlamydia
7	380.2	23.2	1896	6	AX098423	AX098423 Sequence
8	380.2	23.2	2300	8	STP1ATDTR	STP1ATDTR
9	374.8	22.9	2126	8	GS0251356	GS0251356
10	365	22.3	2122	8	AY098893	AY098893 Citrus hy
11	352.4	21.5	2110	8	AF428316	AF428316 Arabidops
12	352.4	21.5	2181	8	ATAMPWR	ATAMPWR
13	350.8	21.4	2151	8	AY045903	AY045903 Arabidops
14	336.6	20.6	1896	8	AY128844	AY128844 Arabidops
15	336.6	20.6	2090	8	AY081350	AY081350 Arabidops
16	336.6	20.6	2146	8	AY084374	AY084374 Arabidops
17	330	20.2	1770	6	AX098422	AX098422 Sequence
18	301.4	18.4	1823	6	AX100499	AX100499 Sequence
19	301.4	18.4	10163	1	AE001646	AE001646 Chlamydia
20	301.4	18.4	10407	1	AE002174	AE002174 Chlamydia
21	301.4	18.4	300550	1	AP002547	AP002547 Chlamydia
22	293.2	17.9	1604	1	RIRATPDTA	RIRATPDTA
23	293.2	17.9	282610	1	RPPXX01	RPPXX01
24	289.4	17.7	14117	1	AE008575	AE008575 Rickettsi
25	283.8	17.3	2139	8	ATATP2	ATATP2
26	274.2	16.8	1505	12	AF087957	AF087957 Synthetic
27	270.2	16.5	1900	1	AY120885	AY120885 Holospora
28	256.4	15.7	24297	1	AE001323	AE001323 Chlamydia
29	252.2	15.4	1623	1	TAJ10587	TAJ10587 Chlamydia
30	232.8	14.2	312430	1	RPPXX02	RPPXX02
31	227.2	13.9	12084	1	AE002346	AE002346 Chlamydia
32	217.2	13.3	10893	1	AE008630	AE008630 Rickettsi
33	215.8	13.2	10200	1	AE008626	AE008626 Rickettsi
34	214	13.1	312430	1	RPPXX02	RPPXX02
35	211.8	12.9	13627	1	AE008614	AE008614 Rickettsi
36	200.8	12.3	279110	1	RPPXX03	RPPXX03
37	199.2	12.2	237523	1	RPPXX04	RPPXX04
38	194.4	11.9	12723	1	AE008663	AE008663 Rickettsi
39	139.8	8.5	132990	8	AC018848	AC018848 Arabidops
40	138	8.4	217975	2	AC126791	AC126791 Medicago
41	136.6	8.3	174617	2	AP003977	AP003977 Oryza sat
42	134.8	8.2	156465	8	AP003234	AP003234 Oryza sat
43	134.8	8.2	172003	8	AP002910	AP002910 Oryza sat
44	132.2	8.1	91001	8	AC013453	AC013453 Arabidops
45	127.8	7.8	148018	2	AP003974	AP003974 Oryza sat

# ALIGNMENTS

RESULT 1  
AE002202/c  
LOCUS  
DEFINITION  
Chlamydia pneumoniae AR39, section 33 of 94 of the complete genome.  
ACCESSION  
AE002202 AE002161  
VERSION  
AE002202.2 GI:8163425  
KEYWORDS  
SOURCE  
ORGANISM  
Chlamydia pneumoniae AR39.  
REFERENCE  
1 (bases 1 to 13389)  
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Unayam,L.A., Uppack,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
2 (bases 1 to 13389)  
Read, T. D., Brunham, R. C., Shen, C., Gill, S. R., Heidelberg, J. F., White, O., Hickey, E. K., Peterson, J., Umayam, L. A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouli, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J., McClary, G., Salzberg, S. L., Eisen, J. and Fraser, C. M.  
Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
10684935  
20150255

TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
1. .13389  
/organism="Chlamydia pneumoniae AR39"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/note="synonym: Chlamydia pneumoniae AR39"  
163. .339  
/gene="CP0400"  
163. .339  
/gene="CP0400"  
/note="hypothetical protein; identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38245.1"  
/db\_xref="GI:7189325"  
/translation="MLSYMKLIPQDIERETDSKRKKITFTIYCSSKVLGHFFS  
HDKHNKTHESIGV"  
597. .1448  
/gene="CP0401"  
597. .1448  
/gene="CP0401"  
/note="hypothetical protein; identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38246.1"  
/db\_xref="GI:7189326"  
/translation="MVNIQPYRNQVNSQATQPSVQCPALSLITVVAVALIYA  
LVQSLSLIEGTALVLSLIFRSAMFMYKMKQEKELIPKIMELIOEHPSI  
VVDIRDOEVSIYIEIHLISILKTNFEDKAPVLOEKLOFGIEKEDVHSPKPNF  
BEILLOHPLHMLGRLYPMVSDVTPGYGYWCGPLGLYENAPSLFERRSLLKRI  
SFGFALLIEDGLKKNWSSSELVQIRQNLFRYYADKEVDEALNDVDEGFDLSLHL  
IFSHKLS"  
1455. .1808  
/gene="CP0402"  
1455. .1808  
/gene="CP0402"  
/note="hypothetical protein; identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38247.1"  
/db\_xref="GI:7189327"  
/translation="MOLFQYNESGMDMLCDPDSGSGFQSLRVGLHSSWALYEAK  
EQFLPEVSLITMELEIMOLLSKPTAHGAKDLCNVFHKHFRQFLGSLDINQRP  
ENTFLNPKTHLDRE"  
1842. .1958  
/gene="CP0403"  
1842. .1958  
/gene="CP0403"  
/note="hypothetical protein; identified by Glimmer2; putative"

gene  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38248.1"  
/db\_xref="GI:7189328"  
/translation="MKSLSVRSAPFSGIVKSLPNLKGKRELFEMGFTTGNs"  
2019. .3326  
/gene="CP0404"  
2019. .3326  
/gene="CP0404"  
/note="hypothetical protein; identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38249.1"  
/db\_xref="GI:7189329"  
/translation="MAMSKTIVPALFNKHCCHSTPFLSRTIVRIASLFCIGAL  
AALGLAPVSVIVSVLAFVILSLVIALIFGEKKLPPTPIIPDRFTYIDEA  
YGLSISAFVREQVTVLAEFRQFSTALCNSIPEEKLOPSELSKYSVFCISRLAGD  
LEKNWPIFEEDLSQTCPLWLOKFRISAGDPOVCRDLVPRECYWGLGYSSTAK  
ATICKETHHTLOQTKREDVLLKRAKIQEKMDPEKAIYERYITTYTAGTILKTEA  
GGLTKETISKELELSLHGSFSDQLITQLPROMDMLCFVDNSTAYNLOALVGA  
LSOVLDESSIDEDVNLGLYIQLDKFAVAFASDSEPKELGLFLRLHSVSKRL  
ESVLRQGLRIALDEHGNARARVYDVNEVTGARIRKTSIFPKD"  
/gene="CP0405"  
complement(3404. .4795)  
/gene="CP0405"  
complement(4476. .5021)  
/note="hypothetical protein; identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38250.1"  
/db\_xref="GI:7189330"  
/translation="MLNLRYNTQIDEPKCMSNTSPYIQNNRSCNYFELKNSTTH  
IVISAAILCGLAIFLCAVAVSVILSGALLGLLALIGVILIGIKKIPMISKEQ  
VEPOLVNRIRAHYPKFVDFVSEAKPMLKOLISFDILNLOHSEVSSNTYVNSEL  
QOKIDFEGIRKKTIEVETASIKLESASASRPPLSLAKTLOKVPFFNIGETSIG  
SKVELHVRKKTIGSLEEDSDYIKPEKLPYVLPIDFRTNSIINLHTLVARVL  
TRDVFOHLKYAALNGEWNLNSDLNTMKQDLFAVYHAAYOSKHLQPSLOEDFYML  
LICIFKHRSWKQMSLKITVPADLMENLCITLDHGRPODMEFASLIGLYDGLLH  
KESEAFSLSLTSLSDQFKTIRQSTNAMPLEMLATHNHSFRLPITVHPLKRSVE  
SQPEDESSLLIG"  
complement(4476. .5021)  
/gene="CP0406"  
complement(4476. .5021)  
/gene="CP0406"  
/note="hypothetical protein; identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38251.1"  
/db\_xref="GI:7189331"  
/translation="MRPRNKKSLIFIKRIYVSSQSGEKGARPFKKSTTSVLILL  
LEATFNPFSSITIONNKNKFNKNISINRIFVKFTI"  
complement(5081. .6361)  
/gene="CP0407"  
complement(5081. .6361)  
/gene="CP0407"  
/note="hypothetical protein; identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38252.1"  
/db\_xref="GI:8163426"  
/translation="MRHSGCTMKCSPTLVPPIFLKNDCEHSCSKIKTTARLLIG  
LVIALVSAISFVLAAPISYVIGGTALAVIILITLVALLAKSKVLPINELKTI  
LYNPKREVEFYFWKTHSLVNEKIFINCKWSGDLPNNHKAKEAFGLDILSIDLT  
LPPEFEILLQNCPLWLSHFIDKTESVAGEIGLNKQKVVYGLGLPLAFHKGYTTTFH



```

SYTRPLTLISQYKFLYSKSNQNDSPVKTCBIEIKELPHNMLFRKDVOGISQ
FLFPEFHGHTWEOAOIMOLINPDNMKMLCOFDRAGRSNATGEGFINTNEDPY
SSNEPYNMTWQIMOLIKVLEKKEKESPHRPSALVOKCVTTHNOLNKRQWVRNT
SSQWTSLSLPQAPHAQYKLEKKEESSLPISLSL"
gene
complement(6479, .8026)
/gene="CP0408"
complement(6479, .8026)
/gene="CP0408"
GB:AJ235269; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ADP, ATP carrier protein"
/protein_id="AAF38252.1"
/db_xref="GI:7189332"
/translation="MTKTEKPEFGKLSFLMPHPIHTELKVLPMFLMEFCTITNYTL
RDYDYLIVGAPSGAEAIPIKFLWVPCAIIPLMLYAKLSNLSKQALYVAAGTPE
LIFPALPPTVYPLRDYLHPEFADRLQAILPGLIGLVALIRMTFAEVLAEWG
SVMLSLMWFGEANEITKHEAKREYALIGANISLASGALVWASKLRASVEGVD
PMGISLRLLMAMTIVSGVLEMAKYMWIKKNVLPDPRFNPENPMQKGAQPKNNMD
SELYARSPYLLIALIVAGICINIEYWKSQLQLOYNNNDYSEPMGNFSFMTG
VVSYLMLFVGNGVYIRKFGWLFGALVTPVNLVLTGIVFELVIFRNQASGLVAMEGTT
PLMLAVVGAIGIONLSKSTKYPALFDSKEMAYIPLDQOKVKGAAADLVVARFGSG
GALIOGGLVYCGSIGAMTPYLAIVLFIILAIWVSATKLKLELAQSALKEQVAVQ
DSAPASS"
gene
complement(8227, .8742)
/gene="CP0409"
complement(8227, .8742)
/gene="CP0409"
/note="conserved hypothetical protein; identified by
glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF38253.1"
/db_xref="GI:7189333"
/translation="MGPMVADEQSHKEIEIKAIKRVCKGNKENDLCRYLPGPSG
YMHFTLKMKMSAPBEOLKMLKTFILSESPRIINPBRAPROSKRRDDINTKTD
IERYLELAROYGDNDLLARFSPKPLTSLKRELLIRISNGIVSELMNAVYEAQKAVS
SPNLEVTSPFY"
gene
complement(8910, .9113)
Query Match          99.8%; Score 1633.8; DB 1; Length 13389;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1635; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 421 AATTTGCTGACCGTTTACAGCCATCTACCCAGAGATTGCTAGACGTCGTTGCCATCT
|||||
DB 7656 AATTGGCTGACCGCTTTACAGCCATCTACCTCCAGAGATTGCTAGACGTCGTTGCCATCT
OY 481 TAAGAACTGACATTTGCTGCAATTTATGTAATCTGCTGMACTATGAGGGAACGCTGATGC
|||||
DB 7596 TAAGAACTGACATTTGCTGCAATTTATGTAATCTGCTGMACTATGAGGGAACGCTGATGC
OY 541 TATCTCTAATGTTCTGGGGAATTTGCTAATGAATTAACAAAATCCACGAGCAAAAGCGTT
|||||
DB 7536 TATCTCTAATGTTCTGGGGAATTTGCTAATGAATTAACAAAATCCACGAGCAAAAGCGTT
OY 601 TCTAGCGCTTTGGGTAACGAGCAATATTTCTTTCTTCTTCTTCTGCTGCTGCAATTTG
|||||
DB 7476 TCTAGCGCTTTGGGTAACGAGCAATATTTCTTTCTTCTTCTTCTGCTGCTGCAATTTG
OY 661 TTGAGGCTTCAAGTGGAGAGCTTCGTTTCTGAAGGTGTAGATCCTTGGGGAATTTCTT
|||||
DB 7416 TTGAGGCTTCAAGTGGAGAGCTTCGTTTCTGAAGGTGTAGATCCTTGGGGAATTTCTT
OY 721 TACGCTTTTGAATGGCTATGACTATTTGATCTGACCTGTCTTATGCGCAATTAAGTGT
|||||
DB 7356 TACGCTTTTGAATGGCTATGACTATTTGATCTGACCTGTCTTATGCGCAATTAAGTGT
OY 781 GGATCAATAGAAAGCATTTGACCGATCCGCTGCTCTATATATCCAGAAAGCAAAAGG
|||||
DB 7296 GGATCAATAGAAAGCATTTGACCGATCCGCTGCTCTATATATCCAGAAAGCAAAAGG
OY 841 GGAAGAAAGGCTTAACCTAAATGAATGAAAGATAGCTCTCTCTCTCTCTCTCTCTCTCTCT
|||||
DB 7236 GGAAGAAAGGCTTAACCTAAATGAATGAAAGATAGCTCTCTCTCTCTCTCTCTCTCTCTCT
OY 901 CTCTCTATATTTCTTTTATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
|||||
DB 7176 CTCTCTATATTTCTTTTATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
OY 961 TCGAGCTGCTGGAAGAAATGCAAGTGAAGCAATGCAATATGCAATATGCAATATGCAATATG
|||||
DB 7116 TCGAGCTGCTGGAAGAAATGCAAGTGAAGCAATGCAATATGCAATATGCAATATGCAATATG
OY 1021 AGTCATGAGGAACTTCTCTCTGACAGCGGTAGTATCCGTACTATCATGATATTTG
|||||
DB 7056 AGTCATGAGGAACTTCTCTCTGACAGCGGTAGTATCCGTACTATCATGATATTTG
OY 1081 TTGTTGTTAGCTCATTCGTAATTTGATGGTTAACTGGAGCCCTAGTACTCTCTGCA
|||||
DB 6996 TTGTTGTTAGCTCATTCGTAATTTGATGGTTAACTGGAGCCCTAGTACTCTCTGCA
OY 1141 TGGTCTCTCTAAGAGTATCGTTTCTGCTCTGCTCTGTTATCTTTAGAAACCAAGCTTCG
|||||
DB 6936 TGGTCTCTCTAAGAGTATCGTTTCTGCTCTGCTCTGTTATCTTTAGAAACCAAGCTTCG
OY 1201 GGGCTGGTGCATTTGTTGCTGTAACGCTCTCATATGCTAGCTGTTGGGAGGATATAC
|||||
DB 6876 GGGCTGGTGCATTTGTTGCTGTAACGCTCTCATATGCTAGCTGTTGGGAGGATATAC
OY 1261 AGAATATTTCTTTGAAATCCACAAAATACGCTCTCTTTGACTCAACTAAGAAATGGCCT
|||||
DB 6816 AGAATATTTCTTTGAAATCCACAAAATACGCTCTCTTTGACTCAACTAAGAAATGGCCT
OY 1321 ATATCCCTCTTGACCAAGAGCAAAAGCAAAAGGTAAGGCTCTATGTAAGTTGGCG
|||||
DB 6756 ATATCCCTCTTGACCAAGAGCAAAAGGTAAGGCTCTATGTAAGTTGGCG
OY 1381 CCCGCTTCGAAATACAGAGAGGCTTTAATCAACAGAGTTGCTCTATCTGTGGA
|||||
DB 6696 CCCGCTTCGAAATACAGAGAGGCTTTAATCAACAGAGTTGCTCTATCTGTGGA
OY 1441 GTATTGAGCTATGACCCCTTATCTTGGAGTGAATCTCTCTCTCTCTCTCTCTCTCTCTCT
|||||
DB 6636 GTATTGAGCTATGACCCCTTATCTTGGAGTGAATCTCTCTCTCTCTCTCTCTCTCTCTCT

```

QY 1501 TGGTTTCGCACTAGTTAAACAACATATTCCTTAGCCGAGTCGCTCTTAAAGACAAG 1560  
|||||  
Db 6576 TGGTTTCGCACTAGTTAAACAACATATTCCTTAGCCGAGTCGCTCTTAAAGACAAG 6517  
|||||  
QY 1561 AAGTGGCAAGAAATTCAGCTCCTGCTCTCATAGAGTTGCTCTTACTCTTGTT 1620  
|||||  
Db 6516 AAGTGGCAAGAAATTCAGCTCCTGCTCTCATAGAGTTGCTCTTACTCTTGTT 6457  
|||||  
QY 1621 GATCCCTACCTGCTTTT 1637  
|||||  
Db 6456 GATCCCTACCTGCTTTT 6440  
|||||  
RESULT 2  
AE001619 11421 bp DNA linear BCT 01-DEC-2000  
LOCUS Chlamydia pneumoniae section 35 of 103 of the complete genome.  
DEFINITION AE001619 AE001363  
ACCESSION AE001619.1 GI:4376620  
VERSION  
KEYWORDS  
SOURCE Chlamydia pneumoniae CML029.  
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE 1 (bases 1 to 11421)  
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,  
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 2 (bases 1 to 11421)  
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,  
Grimwood,J., Davis,R.W. and Stephens,R.S.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University  
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
FEATURES  
source location/Qualifiers  
1. 11421  
/organism="Chlamydia pneumoniae CML029"  
/strain="CML029"  
/db\_xref="taxon:115713"  
complement(95..1069)  
/gene="ytdg"  
/note="Cpn0346"  
complement(95..1069)  
/gene="ytdg"  
/codon\_start=1  
/transl\_table=11  
/product="Integral Membrane Protein"  
/protein\_id="AADI8490.1"  
/db\_xref="GI:4376621"  
/translation="MAIGSPYGVSPFOFVSFRLFGSLFTGSLYTDIQLIVE  
LAISGSAFAGTFLVLRKMAVNAVSHVLFGLVGCLEFTHOITLTLSTGLAAMA  
TAMLTGLIYIRNTEKVSRESSTALVSLFSLSLVLLYPMKNKAITGIELVGNAD  
SLRKEDIFPVTIYILANAVITTFAPRSVLCSSPDSVASSLGIPIRLVYDLIIFQJSA  
CLVGARAVGVNALAFLIIPSLIAKVIASIRSLAMASLVFSIGTAFLAPASSRAIL  
SAYDLGISTGISVFLTMVYIKFISYRGYKSKFEKISSEKSSQY"  
complement(1069..2415)  
/gene="ytdg"  
/note="Cpn0347"  
complement(1069..2415)  
/gene="ytdg"  
/codon\_start=1  
/transl\_table=11  
/product="Integral Membrane Protein"  
/protein\_id="AADI8491.1"  
/db\_xref="GI:4376622"  
/translation="MLSCVFSDFIFLSSFLAVTLICMTALMCTIILISKOPILSLSL  
SHASYPGLVGLAMAOYVSLQASIFMIVLFGCAASFGYGIIVFLGKVKLHKDSLA  
CFVLVEFAIGVILASVRESPTLVNRINAVYGAATLGFLEATLAAVPCASJFA  
LMMVROIYVTFDKDPAVNGIKIVYELSLIFTSIVYSGVRSVGYLISAMRYA  
PSIGARQISRLSTIILSLAFGGISGALCSYISVAFTCAIIGQQAIVPTLPTGPLY  
VICAGLLAGICLLFSRSGWVIRFVRKRHFSFKDDEHLKFWHISHNLNENISVAD

EVCSYKYQYEFKPPFPRMVRVQILEMKGYYKKEDDYRLTKKRSALRLVRAHRLME  
SYLVNSLDFSKSEVHLEAEIHEVLTLELDHTLTLEILNDCYDPHROIIPNKKREV"  
complement(2418..3197)  
/gene="ytdg\_1"  
/note="Cpn0348"  
complement(2418..3197)  
/gene="ytdg\_1"  
/codon\_start=1  
/transl\_table=11  
/product="rRNA methylase"  
/protein\_id="AADI8492.1"  
/db\_xref="GI:4376623"  
/translation="MNVKDEPMSVHNLGVNYEHAVALYHISFSLGKSLTALGPNG  
AGKSTLKLASLIGLKIPSSGTVYEPFNOKFKKVRORIAYMPQASVDMDFPMTVIDLALM  
GCYSYKGMGRISDDDRREAFHILERVGLSDVROJGLSGGQOQRAFLRALMOKA  
DLYIMDELFSADMSFKTSYGVLOELRDGKTIIVYVHDLISHYRQLFDHIVLLNKLRL  
ICCGPTDECLNGDIPFTPTGCEHLELQTLAKLSNGKQFGSC"  
complement(3181..4164)  
/gene="ytdg"  
/note="Cpn0349"  
complement(3181..4164)  
/gene="ytdg"  
/codon\_start=1  
/transl\_table=11  
/product="Solute Protein Binding Family"  
/protein\_id="AADI8493.1"  
/db\_xref="GI:4376624"  
/translation="MDAKMGYIKFVMRIFCFVACGTTFGCTNSGPNOMASRPTISM  
NRNTHOCVEKVRNRLATAVILKIGSLDPHAYEWKGGKDKDIASAVIFCGLEBHTL  
SLRHLNPNNSVKNLGERLJARGAEVPLEDEGJCDPRIMWDLISIMEVAIEIEVLE  
KPEPWSAEFPAENSEIELVCMSILDSMAKQCLSTPENLRYLVSGHNAFSTYRYLAT  
PEEVASGAMRSRCSIEGSLSPDAQISVRDIAVAVDYINEDVSVFVEDTLNDALAK  
IVSLKSLHLVLAQKPLYSNDVNDVYSTFKHNVCILITELGVALCEQR"  
4575..5081  
/gene="Cpn0350"  
4575..5081  
/gene="Cpn0350"  
/codon\_start=1  
/transl\_table=11  
/product="C7066 hypothetical protein"  
/protein\_id="AADI8494.1"  
/db\_xref="GI:4376625"  
/translation="MAVESHIKEIEKLIGKAIKRYCGKKNENDLCRYLPBGSGYWH  
HTLLRKMSAAPQLKMLKFTLSESTPTIIPKRAPRSGSKRRRDIPIFTYDIER  
VLELARQVGDLDLARSPPKPLTSIKRELIRSRINGVISVELMNAVVEAVKAVSSPN  
LEVTSFV"  
5282..6829  
/gene="adt\_1"  
/note="Cpn0351"  
5282..6829  
/gene="adt\_1"  
/codon\_start=1  
/transl\_table=11  
/product="ADP/ATP Translocase"  
/protein\_id="AADI8495.1"  
/db\_xref="GI:4376626"  
/translation="MTKTEKPPGKLSRFLPPIHTHELKVLPMPLMFCITFNVTYL  
RDTKDLIVAGPSGAEAIPKIFMLVIVKPIIEMLIYAKLSLILSKOALFYAVYTL  
LIFPALFPYIYPLRVLHPTERRADRLQALIPGLGIVAILNNMFFAAYIVLAELMG  
SVLSLMEFGFANEITKIHAKRFTALFGIGANISLASERATVMSKLAASVSEGQ  
PWGISLRLMAAMTIVSGLVLMASVWINKVNLNDPRYVNEEQKKGKPKMNKMD  
SFLYLARSPYIILALVIAVYICINLIEVTMSQKLQYPMNDYSEFGNFSFMTG  
VVSVLIMLFVGIVIRKFGMLGALVTPVWVLTGIVEFALVIFRNQASGLVAMFGTT  
PLMLAVVGAIONILSKSTPKYALFDSKTEKAYIPLDOEOKVKKKAIIDVYAARFGSG  
GALIQGLLVIGSGISGAMPYLAIVLLFTIATLVSATRLKMLFLAQSALKEDEVAOE  
DSAPASS"  
6968..8227  
/gene="Cpn0352"  
6968..8227  
/gene="Cpn0352"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"

```
/protein_id="AADI8496.1"
/db_xref="GI:4376627"
/translation="MKCSPLTVPHIFLKNDCBCHRSGLKIRTIARLIGLIVLAVS
ALSPVLAIPISYAIGSTIAAIVIIILITLVALLKSKVLPINPELOKTIYRYEK
EVPFTVTHSLIVNELKIFLIMKSGIDLPENMLKKKAEARIDILKSIDLIPFEE
ILQNCPLVWLMSHFIDKTESVAGEIKNKQKVGGLAFHNGYITIEHSYRPL
TLISESQKFLYSKASKNQMSVSKTCGEIEFEELPHNMIFRKDVQGISQFLF
HGITEAOAMITOLINPDNMKMLCOFDKAGGCHSMATFGGFLNTEINMEDPVSS
VNEWTKELVLEKVESPMHPASALVOKICVNTTHONLKRKMQVRNMTSSQWTS
LPOYAFHAGTYRKLEKIESLPIRSSL"
gene
8287..8532
/gene="Cpno353"
8287..8532
/gene="Cpno353"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8497.1"
/db_xref="GI:4376628"
/translation="MFRNFKSLIFIKIRIYSQSGKEOKGARDPFKKSLTSLVLL
LEAIFENFESSLIQNNFNKNKKNKINIRIFVFTI"
gene
8561..9904
/gene="Cpno354"
8561..9904
/gene="Cpno354"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8498.1"
/db_xref="GI:4376629"
/translation="MSNTSPVIGNNRSCTNYFELKNTSTIHIVISAILLGLALAF
CVAPKSYILSGALIGLILGILALIGVILIPMISKEQVPOEIVNIRAHYK
FVDFSEAKPNLKDILSFIDILNOLSEVSNVNVSELOKIDTIEEIALKNE
VRKSLKRLEBAASRPFLPSLPKILQKVEFFPLGERSKSVYELHRRKTKGSL
EEDLSDYIKPEMLPTWLPIDFRPTNSIUNLHTLVARLVDFDVOHLKVALNGE
WNLNSDLNTHKQOLQFAKHAAYOSYKHLSPQSEDEYVLLICFIRKHSWQMSL
IKVYPADLMENLCCITLDHNGRQDMESALIGITLQGLHKESEAFSLTSLSD
QFKTIRROSTNIAMELEMLATHNSTFSLPITVHPLKRSVFSQPEDESSLIG"
gene
complement(9982..11283)
/gene="Cpno355"
complement(9982..11283)
/gene="Cpno355"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8499.1"
/db_xref="GI:4376630"
/translation="MSMTIVPHALFKNHCHSFPPLSSRTIVRIATSLFCIGALAA
LGCLAPVSYVGSVLAIFVILSVILALIFGEKLPPTPIIPDFTVIDEAYG
LSISAFVREDOVTLAEFRQFTALCNISPEKIKOLPELSKVESGISRAGDLE
KNMPIFEDLISORTCLPYWLOKPLSAGPOVCRDLGVRECYGWTGPGISTAKAT
IFCEETHHITIOQLKEDVLLKNAKLOEKMPDVKAVEITYTYTARQTLKNEAG
LTKETISKELLNLHGYSPDOLITOLPRDAMDLCFVNSTAYNLQDLALGALS
SÖNLDESSIDFVNLGLYIÖDLKEAÖVATASADEPKELGKFLRLHLSVSRLLES
VLROGLHRIALEHGNARARYVDNVFVGARHRTSIFFKD"
BASE COUNT 3791 a 2376 c 1936 g 3318 t
ORIGIN
Query Match 99.7%; Score 1632.2; DB 1; Length 11421;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 181 TACGGATACAAAAGACACTTATTGTGGAGCTGCTGTTCTGTGACAGGCAATAC 240
|||||
DB 5412 TACGGATACAAAAGACACTTATTGTGGAGCTGCTGTTCTGTGACAGGCAATAC 5471
OY 241 CTTTCATCAAGTTTGGCTGTGTGCCCTGTGCTATTATCTTATATATGCAA 300
|||||
DB 5472 CTTTCATCAAGTTTGGCTGTGTGCCCTGTGCTATTATCTTATATATGCAA 5531
OY 301 AGCTAAGTATATTTTAAGTAAGCAGGCTTATTTTATGCACTGGGAAGCCCTTTTAA 360
|||||
DB 5532 AGCTAAGTATATTTTAAGTAAGCAGGCTTATTTTATGCACTGGGAAGCCCTTTTAA 5591
OY 361 TTTTCTTGCCCTGTGCCGACTGTAATTTATCCGCTAGGCAATTTTACATCCCTAACG 420
|||||
DB 5592 TTTTCTTGCCCTGTGCCGACTGTAATTTATCCGCTAGGCAATTTTACATCCCTAACG 5651
OY 421 AATTTGCTGACCGTTTACAGGCCATCTACCTCCAGATTTGCTAGACTGTTGCCATCT 480
|||||
DB 5652 AATTTGCTGACCGTTTACAGGCCATCTACCTCCAGATTTGCTAGACTGTTGCCATCT 5711
OY 481 TAAGAACTGACATTTGCTGCTATTTATCTTCTGTAACATATAGGGAAGCCTATGC 540
|||||
DB 5712 TAAGAACTGACATTTGCTGCTATTTATCTTCTGTAACATATAGGGAAGCCTATGC 5771
OY 541 TATCTCTAATGTTTGGGGATTGGTAATGAATTAACAAAATTCACAGCAAGCAAGCGTT 600
|||||
DB 5772 TATCTCTAATGTTTGGGGATTGGTAATGAATTAACAAAATTCACAGCAAGCAAGCGTT 5831
OY 601 TCTACGCTCTGTTTGGGATTCGAGAGCTAATATTTCTTACTAGCTCTGCTGCAATTG 660
|||||
DB 5832 TCTACGCTCTGTTTGGGATTCGAGAGCTAATATTTCTTACTAGCTCTGCTGCAATTG 5891
OY 661 TTTGGGCTTCAAGAGTTGAGAGCTTCCGTTTCTGAAGGCTGAGATCCTTGGGAATTTCTT 720
|||||
DB 5892 TTTGGGCTTCAAGAGTTGAGAGCTTCCGTTTCTGAAGGCTGAGATCCTTGGGAATTTCTT 5951
OY 721 TACGCTCTGTTTGGGATTCGAGAGCTAATATTTCTTACTAGCTCTGCTGCAATTG 780
|||||
DB 5952 TACGCTCTGTTTGGGATTCGAGAGCTAATATTTCTTACTAGCTCTGCTGCAATTG 6011
OY 781 GGATCAATTAAGAACGTAATTAAGCAGATCTGCTTATTAATCAGAGAAATGCAAAAGC 840
|||||
DB 6012 GGATCAATTAAGAACGTAATTAAGCAGATCTGCTTATTAATCAGAGAAATGCAAAAGC 6071
OY 841 GGAAGAAAGGCTCTAAGCTAATAATGATATGAAGATAGCTTCTCTATCTTGATAGAT 900
|||||
DB 6072 GGAAGAAAGGCTCTAAGCTAATAATGATATGAAGATAGCTTCTCTATCTTGATAGAT 9131
OY 901 CTCCTTAATCTTTTATTAACCTCTGCTGATGCTATGCTATGCTATGCTATTAACCTTAA 960
|||||
DB 6132 CTCCTTAATCTTTTATTAACCTCTGCTGATGCTATGCTATGCTATGCTATTAACCTTAA 6191
OY 961 TCGAAGTGACTGGAAGAGTCAAGTCAAACTCAATATCTCAATATGAATGAATGACTATAGTG 1020
|||||
DB 6192 TCGAAGTGACTGGAAGAGTCAAGTCAAACTCAATATCTCAATATGAATGACTATAGTG 6251
OY 1021 AGTTATGAGGGAACCTTCTCTTCTGAGCTGGCGTATGATCCCTACTTATCATGCTATTTTG 1080
|||||
DB 6252 AGTTATGAGGGAACCTTCTCTTCTGAGCTGGCGTATGATCCCTACTTATCATGCTATTTTG 6311
OY 1081 TTTGGTGTAAGCTCAATGCTTAATTTGATGCTTAACGAGGAGCCCTAGCACTCTCTGCA 1140
|||||
DB 6312 TTTGGTGTAAGCTCAATGCTTAATTTGATGCTTAACGAGGAGCCCTAGCACTCTCTGCA 6371
OY 1141 TGGTCTCTCAACAGGTATGCTTCTTCTGAGCTTGTATCTTTTGAAGCAAGCTTCTG 1200
|||||
DB 6372 TGGTCTCTCAACAGGTATGCTTCTTCTGAGCTTGTATCTTTTGAAGCAAGCTTCTG 6431
OY 1201 GGGTGGTGCTATGTTGGGTACACTCTCTCATCTAGCTGTGTTGTGGAGCTATATAC 1260
|||||
DB 6432 GGGTGGTGCTATGTTGGGTACACTCTCTCATCTAGCTGTGTTGTGGAGCTATATAC 6491
```

QY 1261 AGAATATTCCTTTCGAATTCACAAATAACGCTCTCTTGTACTCAACTAAGAATGCGCT 1320  
|||||  
Db 6492 AGAATATTCCTTTCGAATTCACAAATAACGCTCTCTTGTACTCAACTAAGAATGCGCT 6551  
QY 1321 ATATCCCTCTTGACCAAGCAAAAGTCAAGGTAAAGCTGTCTTATGATGAGTGGCG 1380  
|||||  
Db 6552 ATATCCCTCTTGACCAAGCAAAAGTCAAGGTAAAGCTGTCTTATGATGAGTGGCG 6611  
QY 1381 CCCGCTTGGAATAATCAGAGAGAGCTTTAATCCACAAGGTTTGGCTTATCTGTGAA 1440  
|||||  
Db 6612 CCCGCTTGGAATAATCAGAGAGAGCTTTAATCCACAAGGTTTGGCTTATCTGTGAA 6671  
QY 1441 GTATTGAGCTATGACCCCTTATCTTGAGATGATCTCTTTCATCATGCTATTTGGT 1500  
|||||  
Db 6672 GTATTGAGCTATGACCCCTTATCTTGAGATGATCTCTTTCATCATGCTATTTGGT 6731  
QY 1501 TGGTTTCTGCACTAAGTTAAACAACACTATTTCTTACGCGACTGTCTCTTAAAGACAAG 1560  
|||||  
Db 6732 TGGTTTCTGCACTAAGTTAAACAACACTATTTCTTACGCGACTGTCTCTTAAAGACAAG 6791  
QY 1561 AAGTGCCTCAAGAAGATTACAGCTCTGCTCTTCATAGAGTTGCTTCTTACTCTTGT 1620  
|||||  
Db 6792 AAGTGCCTCAAGAAGATTACAGCTCTGCTCTTCATAGAGTTGCTTCTTACTCTTGT 6851  
QY 1621 GATCCCTACCTGCTTTT 1637  
|||||  
Db 6852 GATCCCTACCTGCTTTT 6868

RESULT 3  
AP002546  
LOCUS  
DEFINITION  
Chlamydia pneumoniae J138 genomic DNA, complete sequence,  
section 2/4.  
ACCESSION  
AB002546  
AB033802  
AB033803  
AB033804  
AB033805  
AB033806  
AB033807  
AB033808  
AB033809  
AB033810  
AB033811  
AB033812  
AB033813  
AB033814  
AB033815  
AB036079  
AB036080  
AB036081  
AB036082  
AB038348  
AB038349  
BA000008  
AP002546.2  
GI:10176692

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Chlamydia pneumoniae J138 (strain:J138) DNA.  
Chlamydia pneumoniae J138  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

Shirai, M., Hirakawa, H., Ouchi, K., Tabuchi, M., Kishi, F., Kimoto, M.,  
Takeuchi, A., Nishida, J., Shibata, K., Fujinaga, R., Yoneda, H.,  
Matsushima, H., Tanaka, C., Furukawa, S., Miura, K., Nakazawa, A.,  
Ishii, K., Shiba, T., Hattori, M., Kuhara, S. and Nakazawa, T.  
Comparison of outer membrane protein genes omp and pmp in the whole  
genome sequences of Chlamydia pneumoniae isolates from Japan and  
the United States  
J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)  
20298986

Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K.,  
Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.  
Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CML029 from USA  
Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
20330349  
3 (bases 1 to 300650)  
Shirai, M.  
Direct Submission  
Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University  
School of Medicine, Department of Microbiology; 1-1-1  
Minamikogushi, Ube, Yamaguchi 755-8505, Japan  
(E-mail:ms Shirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,  
Fax:81-836-22-2415)

On or before Sep 15, 2000 this sequence version replaced  
gi:6172290, gi:6172292, gi:6172294, gi:6172296, gi:6172326,  
gi:6172338, gi:6172330, gi:6172332, gi:6172334, gi:6172336,  
gi:6172338, gi:6172382, gi:6172384, gi:6172386, gi:6172388,  
gi:6172390, gi:6172392, gi:6172394, gi:6174666, gi:6174668,

FEATURES  
source  
1. 300650  
/organism="Chlamydia pneumoniae J138"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym:Chlamydia pneumoniae (strain J138)"  
144..935  
/gene="CPj0267"  
144..935  
/gene="CPj0267"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAA8477.1"  
/db\_xref="GI:8978641"  
/translation="MSIMSLNKTNALNDEPAVCLNAMPDKYIINDRKTPEACTVTL  
VATIMLTGTVIVLLAMSGPGLSVLTIGSTVTLGALFTIGLVKIKSLAMI  
OYKRYFOEVKOKYKPEPSIPKNDVHKLTCLSPDLIESPSPSPVSKRLACG  
VATVIGVTLTIGAVSVFPCGTGYDLALCVGAGCTGALFVGLAGLRTHSLAIGIM  
KLYITTYLSSNLEERNETVADQREINITYLTRECRQKRKALLE"  
937..1343  
/gene="CPj0268"  
937..1343  
/gene="CPj0268"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAA8478.1"  
/db\_xref="GI:8978642"  
/translation="MSQCSSTSTWEMMKSFVPMKMPPTPLSPISEDEFILAEV  
FVLKPTPENAOANPCTSTPNVENGIDLPILGQNEQNANNPCTSGSNPTSLPA  
PERLPTEENSQEEPEQSONNEDIG"  
complement(1495..2472)  
/gene="CPj0269"  
complement(1495..2472)  
/gene="CPj0269"  
/codon\_start=1  
/transl\_table=1  
/product="dipeptidase"  
/protein\_id="BAA8479.1"  
/db\_xref="GI:8978643"  
/translation="MTIDMHCDSLHPHRCRDPVAVRCSPEQLSGVROQCAIYV  
HSRGEPCDKONLSLFFSLPNQYDGLGSLSEENSGSSQKSLISRTSEMSALMEL  
DTPAPICGLTKLHITKQGLAVIGLWKGDNRRGCTEAPKLSNDGKVLVAKHNP  
GVPLDLSHCSDKLAEDLDYTDADKPLALYASISNRSYLDHRRNIYDHAKEIYR  
KGYTGLNLSVYGDSDGLEKHLHAEVGLISSTVLGSDFFYANDEDFEENCS  
AEAHPIVNOIHRIFSGKAESILSRFAEKFLLQVIVEQVNPKITDVKL"  
complement(2487..3347)  
/gene="ywlC"  
complement(2487..3347)  
/gene="ywlC"  
/codon\_start=1  
/transl\_table=1  
/product="SuA5 superfamily-related protein"  
/protein\_id="BAA8480.1"  
/db\_xref="GI:8978644"  
/translation="MPDKKAQITPSLEPVMSAIHQKIVALLPTDVTYGVFLSLXASA  
EERIVALKDREPSKAFALVNSIEDLENISGTPSLPRAKLAOLFPGATLIVKHNRP  
RFPKRTLAIRIVDSVVRREIVDHCGLTISANISEPSALTAQELFADPADHDLICF  
DGPCHGLESTVAVSDPLTYRGLTISRVIENIAGTEAKIFHRTSHAFSKHKIYIV  
KNOELVSLFSGSLDFKGVGCHEPKPKNFYTRLEALKKTKPSIVFIYDINTSDPEL  
FPLSPYYIR"  
3656..4387  
/gene="CPj0271"  
3656..4387  
/gene="CPj0271"  
/codon\_start=1  
/transl\_table=1  
/product="lysophospholipase esterase"  
/protein\_id="BAA8481.1"  
/db\_xref="GI:8978645"

```

gene
CDS
    /translation="MTDYSFERRKIGNIEAIECPNPDPPIIILCHGYSGLADNLTF
    PSICSFSLRPTWIFPNGLIPLLENDPFSRACPLNVLLESLRYLNSQNYLQEKY
    DELFDVLETKALEELLINLNPENYIIIGFSOGALILATHVLTSQNYAGALIL
    AGARLNGNGEGLKCOQVPEFLOSHGDEBELPYHGHAILNDLLRLNQPFSFHG
    GHEIPSVFOKMOVTPNWDIPARQ"
    /complement(4359..5255)
    /gene="dnaX_2"
    /complement(4359..5255)
    /gene="dnaX_2"
    /codon_start=1
    /transl_table=11
    /product="DNA polymerase III gamma and tau"
    /protein_id="BA98482.1"
    /db_xref="GI:8978646"
    /translation="MHLEENOGWEALLRKVYHDEVPAIILHGFTPLVDKAROLA
    SEILSSPSGEHKYSOKIHPDIYOFPEPGRHLSIDLPGRICKQIYISPEANIKI
    YIIHBDMMTLAISAFLKVEEPKHAVALITLTKAVORLEKTIISRLSLFIEGK
    ILCSKEFTSYLFERYAOCEIPLYEVSQIIKESSEDKOVLROKVOFEMVLELVDYR
    TLNLGLKASALNPEHVEKEIILQLPLLPDKVLLIVESACRSLNNSSAAVLEWAIO
    LVSLOYKEKELVSPGODLSN"
    /complement(5246..5866)
    /gene="tdk"
    /complement(5246..5866)
    /codon_start=1
    /transl_table=11
    /product="chymidylate kinase"
    /protein_id="BA98483.1"
    /db_xref="GI:8978647"
    /translation="MFIVIEGEGSGSKSLAKALGDOLVAODRKVLLREFGGLIGE
    RLRLDLEPHELESCCELEFLGSRQHOEYIIPALRGYIYICRPHDSTIVYG
    GIAGLGADPVADLCSKYVGPPLPLPNVLDDIPADIGLGRKHQKFDKFEKKPLS
    YHNRIRGEFLSLASDPSRYVLVDARESLALIDKVMILHLOGLCT"
    /complement(5871..8375)
    /gene="gyrA_1"
    /complement(5871..8375)
    /gene="gyrA_1"
    /codon_start=1
    /transl_table=11
    /product="DNA gyrase subunit A"
    /protein_id="BA98484.1"
    /db_xref="GI:8978648"
    /translation="MFNKDEIIVPKNEEMKESYLRSMSVYISRALPDIRDGLPS
    ORVLYAMKOLSLSPGAKHKCAKICGDTSGDYHSEVPTPLVRAOWMMAMKYP
    VDGONGESIDGDEPAARMYEARLTHSAMLMDLDQDIYVNYETHEPVPVP
    SKPRLICNCSGGIAPVAMATNIPRNLGELLATLILANQASVDELQVMPGDP
    TGGIICGSEGRSTTTTGKIKIVARLHVEDNEKHESIITTEPYNVKSRLIEO
    IANLVNKTLAGISDVROSDKDGIRVLEIKGSSSEIINRLKFTDYOVTEGAM
    LADKNLPTMSIHMIISAMIRHREKVIIRRTRELNKAETRAHYLEGYLALSCDA
    LVKTIRESGNEHAKERIIESGFTEPOLALILELRYOLGLEAEKIOKEYEELNK
    IAYKOVYSDGLVKDIIIRNELODLKHKVARRTTIEFDADDIRDIEDITNESVLI
    TISGDYKKRMPVYKFEQRRGGHGVTFDKKAGFLKAYSATTKYLLIFTEGO
    CYLKWOLPGEERRAKRPIINLEGIPEBELAAILINFPDAGLFLATKRGV
    KRVSLDASNPBKRGIRALEIDEGELIAACHIVSDEKVMLEPHLGAVERPHEKRV
    PMGRFARGVSVSLKNEEDKYVSCQIVTNSGVILVCDQGGKRSVLDFEFTNGVY
    GVRSLIINERNGVNLGAIPVTDHDSIILMSOGQIRINMDQVRVMSSTOGVRLVH
    KEGDAIVMEKISMENDEVLVSGSEECSDPTSLR"
    /complement(8391..10808)
    /gene="gyrB_1"
    /complement(8391..10808)
    /gene="gyrB_1"
    /codon_start=1
    /transl_table=11
    /product="DNA gyrase subunit B"
    /protein_id="BA98485.1"
    /db_xref="GI:8978649"
    /translation="MDPKENYDASATVLEGLQAVRERPGMIGDYGITGLHMLVE
    VVNSIDAMAGYCSRIDVLELDGIVIVNGRGIPIEVEHERESAKGREGVSALEV
    LTVLHAGSKPKDSYKVGSGLHGVSVCNLSLEKLVATVFKDKCKQYMEFSRGIPV
    PLOYVVSVDROGTETVYPPDKIPSTCTCEVNSIIMRLRELAFNLRGITVIFEDDRD
    SPDKVTFYEGGIGSFVSYLMONKESLSELYICGTRVGDGELTEFFPALQOMNGIS
    ELVSYANNIPTROGTHLTGFTALFVIVITYIKAHNLANNKILAGDEIDRELTLA
    VISVKNPNDPEOGTOKIKIGNSDVSSVAQOVVVEGALTIFPEENPIIAMIIVDKFVA

```

```

OAREAKKARELILKRSALDSARLFGKLIIDLEKPEKCEMYIEVDSAGSAGKGRD
RFOALILSRKILINVEKARLOKIPONOEIGIITIALCGAGDNPENSKLRYRILII
MPDAVDOSLRIITLLIFEFYHVALNLEINCEVYANPOLVYKQKDDRYLISEKEMD
SYLLMLGTNESIILFKSTERELRGALSEFINVILIDVSEFINTELEKAIIPSELEMY
KEGIGYPLYLAPATMGOGGRYLYSDEKESALAOEETHKRIIELYVAAVVDIÖNÖ
LKEYGLIDISYLIPOKNEIVIGNEIDSPSCNCSYCLEEIVNKLKMGKGIEIORYKG
LGEMNADOLMDTMMPEQORTLIHVSILKDAVBADHIFPMLMEVEVPPREFESHLSLI
RIINMDI"
    /complement(10812..11159)
    /gene="CpJ0276"
    /complement(10812..11159)
    /gene="CpJ0276"
    /gene="CpJ0276"
Query Match          99.7%: Score 1632.2; DB 1; Length 300650;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
    1  GAAATAAAAAACTATCAGATAGAAAAATAAAGTATTCAGAGGTAATATGACAAAA 60
    93946  GAAATAAAAAACTATCAGATAGAAAAATAAAGTATTCAGAGGTAATATGACAAAA 94005
    61  CCGAAGAAAAACCTTTTGGAAAAATTCGCTCTTTCTTGCGCCGATACATCTACGAGC 120
    94006  CCGAAGAAAAACCTTTTGGAAAAATTCGCTCTTTCTTGCGCCGATACATCTACGAGC 94065
    121  TAAAGAAAGTTCGCCAATGTTCCTTAATGTCTCTGCTATTCATTAATTAACGCTGT 180
    94066  TAAAGAAAGTTCGCCAATGTTCCTTAATGTCTCTGCTATTCATTAATTAACGCTGT 94125
    181  TACGCGATCAAAAGACACTTATATGTGGAGCTCCTGTTCTGTGACAGAGCAATAC 240
    94126  TACGCGATCAAAAGACACTTATATGTGGAGCTCCTGTTCTGTGACAGAGCAATAC 94185
    241  CTTTCATCAAGTTTGGCTTTGTTGCCCTGTCGTATTAATCTTATGTTATATGCAA 300
    94186  CTTTCATCAAGTTTGGCTTTGTTGCCCTGTCGTATTAATCTTATGTTATATGCAA 94245
    301  ACCTAAGTAAATATTTTAACTAAGACAGGCTTATTTATGACAGGGAAGCCCTTTTAA 360
    94246  ACCTAAGTAAATATTTTAACTAAGACAGGCTTATTTATGACAGGGAAGCCCTTTTAA 94305
    361  TTTTCTTTGCTGCTTCCGACGTAAATTTATCCGCTAGCGATGTTTACATCTACAG 420
    94306  TTTTCTTTGCTGCTTCCGACGTAAATTTATCCGCTAGCGATGTTTACATCTACAG 94365
    421  AATTGCTACCGTTTACAGCCCATCTACCTCCAGAGATTGCTAGAGACTGTTGCCATCT 480
    94366  AATTGCTACCGTTTACAGCCCATCTACCTCCAGAGATTGCTAGAGACTGTTGCCATCT 94425
    481  TAAGAAACGAGCAATTTGCTGCAATTTATGTAAGTCTGCAAGCTATGAGGAGGTCATGC 540
    94426  TAAGAAACGAGCAATTTGCTGCAATTTATGTAAGTCTGCAAGCTATGAGGAGGTCATGC 94485
    541  TATCTCTAATGTTCTGGGATTTGCTAATGAATATCAAAAAATCCAGAGCAAAACGTT 600
    94486  TATCTCTAATGTTCTGGGATTTGCTAATGAATATCAAAAAATCCAGAGCAAAACGTT 94545
    601  TCTACGCTCTTTTCCGTATCGAGCTAAATATTTCTTACTAGCTTGTGCTGCAATTG 660
    94546  TCTACGCTCTTTTCCGTATCGAGCTAAATATTTCTTACTAGCTTGTGCTGCAATTG 94605
    661  TTTGGGCTCAAAAGTGAAGCTGCTGTTGGAAGGTTAGATCCCTGGGAATTTCTT 720
    94606  TTTGGGCTCAAAAGTGAAGCTGCTGTTGGAAGGTTAGATCCCTGGGAATTTCTT 94665
    721  TACGCTTTTGAATGCTATGATTTGATGCACTTGTCTTATGCGCACTTACTGCT 780
    94666  TACGCTTTTGAATGCTATGATTTGATGCACTTGTCTTATGCGCACTTACTGCT 94725
    781  GGATCAATAGAAAGCTATTTGACCGATCTCGCTTCTATATCCAGAGAAATGCAAAAG 840
    94726  GGATCAATAGAAAGCTATTTGACCGATCTCGCTTCTATATCCAGAGAAATGCAAAAG 94785

```

QY 841 GGAAGAAAGGCTCTAAACCTAAATGATATGAAAGATAGCTTCTCTATCTTGATAGAT 900  
 Db 94786 GGAAGAAAGGCTCTAAACCTAAATGATATGAAAGATAGCTTCTCTATCTTGATAGAT 94845  
 QY 901 CTCCTATATCTCTTATTAATCTCTCTGTTATGTCGATGATGATGATGATGATGAT 960  
 Db 94846 CTCCTATATCTCTTATTAATCTCTCTGTTATGTCGATGATGATGATGATGATGAT 94905  
 QY 961 TCGAGTACCTGGAAAGTACAGTGAAGTCAATATCTAATATGATGATGATGATGATGAT 1020  
 Db 94906 TCGAGTACCTGGAAAGTACAGTGAAGTCAATATCTAATATGATGATGATGATGATGAT 94965  
 QY 1021 AGTTCATGAGGACCTCTCTCTCTGATGATGATGATGATGATGATGATGATGATGAT 1080  
 Db 94966 AGTTCATGAGGACCTCTCTCTCTGATGATGATGATGATGATGATGATGATGATGAT 95025  
 QY 1081 TTGGTGTACAGTCACTCTGTAATTTGATGATGATGATGATGATGATGATGATGATGAT 1140  
 Db 95026 TTGGTGTACAGTCACTCTGTAATTTGATGATGATGATGATGATGATGATGATGATGAT 95085  
 QY 1141 TGGTGTCTCTAACAGGTATCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200  
 Db 95086 TGGTGTCTCTAACAGGTATCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 95145  
 QY 1201 GCGTGTGCTATGTTCTGCTACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260  
 Db 95146 GCGTGTGCTATGTTCTGCTACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 95205  
 QY 1261 AGAATATCTCTTGCAGAAATCCAGAAATATGAGCTCTCTCTCTCTCTCTCTCTCTCT 1320  
 Db 95206 AGAATATCTCTTGCAGAAATCCAGAAATATGAGCTCTCTCTCTCTCTCTCTCTCTCT 95265  
 QY 1321 ATATCCCTCTTGACCAAGAGCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 1380  
 Db 95266 ATATCCCTCTTGACCAAGAGCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 95325  
 QY 1381 CCCCTCTGGAAGTCAAGAGAGAGCTTTAATCAAGAGTCTCTCTCTCTCTCTCTCTCT 1440  
 Db 95326 CCCCTCTGGAAGTCAAGAGAGAGCTTTAATCAAGAGTCTCTCTCTCTCTCTCTCTCT 95385  
 QY 1441 GTATTGAGATGACCCCTTATCTGAGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500  
 Db 95386 GTATTGAGATGACCCCTTATCTGAGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 95445  
 QY 1501 TGGTCTCTGCACTAAGTAAACAACATCTTCAAGCAGCTGCTCTTAAAGAACAG 1560  
 Db 95446 TGGTCTCTGCACTAAGTAAACAACATCTTCAAGCAGCTGCTCTTAAAGAACAG 95505  
 QY 1561 AAGTGGCTCAAGAGATTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620  
 Db 95506 AAGTGGCTCAAGAGATTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 95565  
 QY 1621 GATCCCTACCTGCTTTT 1637  
 Db 95566 GATCCCTACCTGCTTTT 95582

RESULT 4  
 AE001281/c 14168 bp DNA linear BCT 30-OCT-2000  
 LOCUS Chlamydia trachomatis section 8 of 87 of the complete genome.  
 DEFINITION Chlamydia trachomatis section 8 of 87 of the complete genome.  
 ACCESSION AE001281 AE001273  
 VERSION AE001281.1 GI:3328454  
 KEYWORDS  
 SOURCE Chlamydia trachomatis.  
 ORGANISM Chlamydia trachomatis.  
 Bacteria; Chlamydiales; Chlamydiales; Chlamydia.  
 REFERENCE  
 1 (bases 1 to 14168)  
 Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,  
 Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
 Koonin, E.V. and Davis, R.W.  
 TITLE Genome sequence of an obligate intracellular pathogen of humans:  
 Chlamydia trachomatis

JOURNAL Science 282 (5389), 754-759 (1998)  
 MEDLINE 9900809  
 PUBMED 9784136  
 REFERENCE 2 (bases 1 to 14168)  
 AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,  
 Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
 MEDLINE 9920606  
 PUBMED 10192388  
 REFERENCE 3 (bases 1 to 14168)  
 AUTHORS Stephens, R.S., Kaiman, S., Lammel, C.J., Fan, J., Marathe, R.,  
 Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
 Koonin, E.V. and Davis, R.W.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University  
 of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA

## FEATURES

source  
 1. 14168  
 /organism="Chlamydia trachomatis"  
 /strain="D/DW-3/CX"  
 /db\_xref="taxon:813"  
 86. 847  
 /gene="flaA"  
 /note="CT061"  
 86. 847  
 /gene="flaA"  
 /codon\_start=1  
 /transl\_table=1  
 /product="Sigma-28/Whig Family"  
 /protein\_id="AAC67652.1"  
 /db\_xref="GI:3328454"  
 /translation="MKTHDLADIMQLYSTKEIHHRDVLIESYLPKVNARLASGM  
 PSYKMDLALGVEGLRAVERDPEKRSFESYAFITKAITIGLRKQDWPVS  
 YQANRLADMDLSRLTGKEDGVDYIKKKAISTLTALALDEKQVNAVLYYDOL  
 VKKIGKILGVESRVSQHSKALKLRGLSSL"  
 1006. 2244  
 /gene="lysS"  
 /note="CT062"  
 1006. 2244  
 /gene="lysS"  
 /codon\_start=1  
 /transl\_table=1  
 /product="tyrosyl tRNA synthetase"  
 /protein\_id="AAC67653.1"  
 /db\_xref="GI:3328456"  
 /translation="MOGLDNLKRGITLSSAGLESYTPVSNYLGFPPPSLHIG  
 HWIGICFLRLAAYGITPVAVLGGATGEMIDPSGYSERSLDDQAVLDNSKRTAAL  
 ASYLPGRIVNADWLGSLMVDPLRDVGNHFRIGMLADVVKORYSDEGISTYEF  
 SYLLGSYDFAHLFEKHNVLVOCGSDQMCNITSGIDYIRRGIGQAYGLYPLIDS  
 KGRKIGTSEGTILMDPLRPYELPQYFLRLPDQESKVMRTLLDNEITFALDER  
 LTPDQAVKRYVIAEIVKYDVHSGEGLAQAQAATSPASKSKSTTETELVAIVESCG  
 VKVARADLGRKMDIYVELGFCSSRGQARRLLIQQRGLYINQELADEQSILDGTOLC  
 FDRYVLDSQGRKKQYIDLN"  
 2254. 3696  
 /gene="gnd"  
 /note="CT063"  
 2254. 3696  
 /gene="gnd"  
 /codon\_start=1  
 /transl\_table=1  
 /product="6-Phosphogluconate Dehydrogenase"  
 /protein\_id="AAC67654.1"  
 /db\_xref="GI:3328457"  
 /translation="MAPNTDILGLIAGLVKNGKILVLMNWDHGSVSYVNSPAKTEPL  
 KDHESGALOGFTTQIEVOSLKRPRKIMIMIRAGAPVDEMIASLPLFEEGLILD  
 KNSYLYDSEQRVYDLKKEGILFVGMVSGGSEBAGRPSIMPGNDAMPALPIQS  
 IAAOVORPCCSWIGTGAGHFAKAVANGIEYDIOLICEYIILKRLNLSLEIGN  
 IFFENQOTDINSYLIGAAALVIAKDNENAIATSTIDVAGCKGTGRVAEDAKICG  
 PMSLIESYARLSTWKEVTRAADEPFIPLCCPPQDEASFIEDVREBALYAAKI  
 STAGFMRLKQVSQDGMWDLNIGELALIMWGGCTIQSAFLDKIHQGFENSPHSLIL  
 ODYKRVLPDSETFRRRAVLAIGSGVAIPCLSSALSFYDGYRTVDSILFVGLRDI

## CDS

gene  
 2254. 3696  
 /gene="gnd"  
 /note="CT063"  
 2254. 3696  
 /gene="gnd"  
 /codon\_start=1  
 /transl\_table=1  
 /product="6-Phosphogluconate Dehydrogenase"  
 /protein\_id="AAC67654.1"  
 /db\_xref="GI:3328457"  
 /translation="MAPNTDILGLIAGLVKNGKILVLMNWDHGSVSYVNSPAKTEPL  
 KDHESGALOGFTTQIEVOSLKRPRKIMIMIRAGAPVDEMIASLPLFEEGLILD  
 KNSYLYDSEQRVYDLKKEGILFVGMVSGGSEBAGRPSIMPGNDAMPALPIQS  
 IAAOVORPCCSWIGTGAGHFAKAVANGIEYDIOLICEYIILKRLNLSLEIGN  
 IFFENQOTDINSYLIGAAALVIAKDNENAIATSTIDVAGCKGTGRVAEDAKICG  
 PMSLIESYARLSTWKEVTRAADEPFIPLCCPPQDEASFIEDVREBALYAAKI  
 STAGFMRLKQVSQDGMWDLNIGELALIMWGGCTIQSAFLDKIHQGFENSPHSLIL  
 ODYKRVLPDSETFRRRAVLAIGSGVAIPCLSSALSFYDGYRTVDSILFVGLRDI

CDS  
 2254. 3696  
 /gene="gnd"  
 /note="CT063"  
 2254. 3696  
 /gene="gnd"  
 /codon\_start=1  
 /transl\_table=1  
 /product="6-Phosphogluconate Dehydrogenase"  
 /protein\_id="AAC67654.1"  
 /db\_xref="GI:3328457"  
 /translation="MAPNTDILGLIAGLVKNGKILVLMNWDHGSVSYVNSPAKTEPL  
 KDHESGALOGFTTQIEVOSLKRPRKIMIMIRAGAPVDEMIASLPLFEEGLILD  
 KNSYLYDSEQRVYDLKKEGILFVGMVSGGSEBAGRPSIMPGNDAMPALPIQS  
 IAAOVORPCCSWIGTGAGHFAKAVANGIEYDIOLICEYIILKRLNLSLEIGN  
 IFFENQOTDINSYLIGAAALVIAKDNENAIATSTIDVAGCKGTGRVAEDAKICG  
 PMSLIESYARLSTWKEVTRAADEPFIPLCCPPQDEASFIEDVREBALYAAKI  
 STAGFMRLKQVSQDGMWDLNIGELALIMWGGCTIQSAFLDKIHQGFENSPHSLIL  
 ODYKRVLPDSETFRRRAVLAIGSGVAIPCLSSALSFYDGYRTVDSILFVGLRDI





QY	420	GAATTGGTGCACCTTTACAGAGGCATCCATCCCTCCAGAAATGCTAGACCTGCTGCATC	479
Db	6598	GCTTTCGCTGTACTCTACATCATCATCTTCTCTCCGGATTATGGATTCACTTGCATG	6909
QY	480	TTAAGAACTGGACATTTGCTGCATTTTATGTACTTGCTGAACCTATGGGAAGCCTCATG	539
Db	6908	CTACGCACTGACATTTGCTGTGTCTACGTACTTTCTGAACCTTTGGGAAGCCTTATG	6849
QY	540	CTATCTCAATGTTCTGGGCATTTGCTAATGAATTTACAATAATCCACGAACAAGCT	599
Db	6848	CTCTCTTGGATGTCTGGGGCTTTGGCCAAAGAATTACTAAATTAAGCAACCTAAAGT	6789
QY	600	TTTACAGCTCTTTCCGCTATCCGGAGCAATATATTCTTCTAGCTCTGCTGGTGCATAT	659
Db	6788	TTTACAGCTCTTTCCGAGTAGAGAGCTAACGTAGCTCTCTTAATTTCTGGTCCAGCTAT	6728
QY	660	GTTTGGCCTTCAAGTTGAGAGCCTCCGTTTCTGGAAGGTATAGATCTTTGGGGAATTTCT	719
Db	6728	ATCTGCTCTTCTAAATTACGCGTAGCGCTGGGGAAGCGGTTGACCCATGGGAGCTACAG	6668
QY	720	TTTACGCTCTTTGATGCTATGACTATTGTACTGGAAGCTTGTCTTATGGCCAGTTACTGG	779
Db	6668	CTCTATTTCCAAAGGCTATGTTCTTGTTCTTGGCGATTAATATGCTCTGTACATGG	6609
QY	780	TGATCAATAGAAAGATATGACCAATCCGCGTTCTATATATCCAGAAATATGCAAG	839
Db	6608	TGATATGACCGCTACGCTTACGCTTACAGATCCAGATTTCTACAAATCTGCAAGCTCAAACT	6549
QY	840	GGGAAAAAGGTGCTTAACCTAAATATGAATATGAAGATAGCTTCTCTATCTTATGATA	899
Db	6548	AAGAAA-----TCTAAGCTAAGATGAGATGGGGGAAGCTTCTTATCTGTATAGA	6495
QY	900	TCTCCTTATATCTTTTATTTAACTCTCTGTGGTTATTGCTATGTATTTGGCTTAATCTTA	959
Db	6494	TCTCCTTATATGCTTCTTACGCTCTTCACTAGTATTTGCTACGGAATCTGCAATTAACTC	6435
QY	960	ATGCAAGTACTGGAAAACTGACGCTGAACCTGCATATCCATATATGAATAGCTATAGT	1019
Db	6434	GTTTAAATTAATCTGGAAAAACCACTCAAAATGCAATTTCCATTAATCAATACATACAGC	6375
QY	1020	GAGTTCATAGGGGAACCTCTCTCTGAGAGCGGTATATCCGCTACTATCATGCTATTT	1079
Db	6374	GCATTTATGGGGAACCTCTCTTCTGACAGAGAGTTGTATCTGTATTTGTATAGCTCTTTC	6315
QY	1080	GTTTGGGTAAAGCTCATTCGTAAATTTGGATGGTTAACTGAGCCCTATGTCACCTCTGTC	1139
Db	6314	ATCGGTGTAAAGCTTATCCCTAGATTCGGGTGATTAACGAGCTTTGGTTAACACTATTC	6255
QY	1140	ATGCTTCTCTACAGGTATCGTTTCTTCGCTCTGTATCTTATCTTTAGAACCAGCTCTT	1199
Db	6254	ATGTTTCTCTGATACAGAGAGCTGTCTTCTTCTGCTCTGTATTTCTTAAGATCATCATCA	6195
QY	1200	GGGCTGTGCTATGTTCTGGTATCACTCTCTCATGCTAGCTGTGGTGTGGAGCTATA	1259
Db	6194	GGTTATGTTCTGTCTTTGGSAACACACCACTAATCTAGCCGTGTGTGTGGAGCCATT	6135
QY	1260	CAGAAATATCTTTGGAATTCACAAATACGCTCTCTTTGACTCACTTAAGAAATGGCC	1319
Db	6134	CAAAATATCTTTCAAAATTAATTAATATGCTCTCTTTGATGCAACATAAGATAGCT	6075
QY	1320	TATATCCCTCTTACCAAGGCAAAAAGTCAAAAGGTAAAGGTGCTATTTGATGTAGTGGC	1379
Db	6074	TATATCCCTATGAGACCAAGGCAAAAAGTAAAGGAAAGCTGATTAAGAGCTTTGGTCC	6015
QY	1380	GCCGCTTGGGAAATACAGAGAGCTTTAATCCAAAGGTTTCTGCTGATATGCTGGA	1439
Db	6014	GCTAGATTCGGTAAATCTGGGGATCTTTGATGATCCAAAGGCTCTCTTGTGCTTTGGCGA	5955
QY	1440	AGTATTTGAGCTATGACCCCTTATCTTGCAGTATCTTCTTTTCAATGCTATTGGT	1499
Db	5954	AGATGCGGGCTATGACCTCATCTTCTAGCTGTGTGCTCTTTGGCAATTAATCATGTGTGG	5895
QY	1500	TTGGTTTCTGCACTAAGTTAAACAACATATTTCTTAAGCGAGTCTGCTTTAAAGAACAA	1559

Dd	5894	TTCACATCTCAGCAACTAAACAACACTGTTTCTTGCGTCATCCGCTGCTTAAGAACAG	5835
Oy	1560	GAAGTGGCTCAAGAGATTCCAG	1582
Dd	5834	GAATTAGCAGAGAGCTGCACGAGC	5812
RESULT 5 LOCUS TAJ10586			
DEFINITION	TAJ10586	1587 bp	DNA linear BCT 23-MAR-1999
ACCESSION	AJ010586		Chlamydia trachomatis nptI gene.
VERSION	AJ010586.1	GI:4538859	
KEYWORDS	nptI gene; nucleoside triphosphate transport protein 1.		
SOURCE	Chlamydia trachomatis.		
ORGANISM	Chlamydia trachomatis		
REFERENCE	Bacteria; Chlamydiaceae; Chlamydia.		
AUTHORS	1 (bases 1 to 1587) Tjaden,J., Winkler,H.H., Schwoppe,C., Van Der Laan,M., Mollmann,T. and Neuhaus,H.E. Two nucleotide transport proteins in Chlamydia trachomatis, one for net nucleoside triphosphate uptake and the other for transport of energy		
JOURNAL	J. Bacteriol.	181 (4),	1196-1202 (1999)
MEDLINE	99138740		
PUBMED	9973346		
REFERENCE	2 (bases 1 to 1587)		
AUTHORS	Neuhaus,E.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-AUG-1998) Neuhaus E., University of Osnabrueck, Biology / Plant Physiology, Barbarastr.11, D-49069 Osnabrueck, FRG		
FEATURES			
SOURCE	Location/Qualifiers 1..1587 <code>/organism=</code> "Chlamydia trachomatis" <code>/db_xref=</code> "taxon:813" 1..1587 <code>/gene=</code> "nptI" 1..1587 <code>/gene=</code> "nptI" <code>/function=</code> "Atp/Adp transport protein" <code>/codon_start=</code> 1 <code>/evidence-experimentl</code> <code>/transl_table=</code> 11 <code>/product=</code> "nucleoside triphosphate transport protein 1" <code>/protein_id=</code> "CAF39534.1" <code>/db_xref=</code> "GI:4538860"		
gene			
CDS			
ORIGIN			
BASE COUNT	384 a	357 c	330 g     516 t
Query Match	52.9%; Score 865.2;	DB 1:	Length 1587; Best Local Similarity 73.3%; Pred. No. 4.3e-196; Matches 1123; Conservativity 0; Mismatches 403; Indels 6; Gaps 1;
Oy	51	ATGAAAAAACGAAACCTTTGGAAAATAATGGCGTCTTTCTTGCGGCCGATACAT	110
Dd	1	ATGACTCAAACCGCGGAAAAACCTTTGGAAAATATGGCGTCTTTCTTGCGCGCATACAC	60
Oy	111	ACTCAGAGACTAAAGAAATTCTGCCAATGTTCCTTAAGTTCTTCTGATTACATTAAAC	170
Dd	61	ATGCATGAGCGTAGAAGAAAGTTCTGCCAATGTTCCTTAATGTCTTCTGATTACATTCAAT	120
Oy	171	TATACGGTGTTAACGGATATACAAAGACACTCTTATTGTGGAGACCTCGTTCCTGTCGCA	230



D	b	1195	GGAGCATTCAAAATATTCTTTCCAAATCAACATAATATGCTCTTGTGAATGCACAATAA	1254
O	y	1311	GAATAGCCTATATTCCTCTTGACCAGAGCAAAAAAGTCAAGGTAAAGCTATTGAT	1370
D	b	1255	GAGCGCTTACATCCCATTGTGATCAAGAGCAAAAAGTTAAAGGGAAGCTGATTGAC	1314
O	y	1371	GTAGTTGCCGCCGCTTGCGSAAATCAGAGAGCTTTTAATCCAACAGGTTTGCTCT	1430
D	b	1315	GTGTGTGCGGTAGATWTCGTAATCTGGGGGATCTTTGATTCACAAGATCTTGTGTC	1374
O	y	1431	ATCTGTGAAGTATTTGAGACTATGACCCCTTATCTGTGAGTGAATCTCTTTCATCTT	1490
D	b	1375	GTTTTCCGGAAGCATCGCGGCTATGACTTCATTCCTAGCTGTGTCTCTTTTGCAATTATC	1434
O	y	1551	AAAGACACAGAAGTGTGCTCAAGAGATTTCAGC	1582
D	b	1495	AAAGAACGAGAAATTAGCAGAAAGCTGCACAGC	1526

RESULT 6  
AE002301/c

LOCUS	AE002301	23773 bp	DNA	linear	BCT 26-MAY-2000
DEFINITION	Chlamydia muridarum, section 33 of 85 of the complete genome.				
ACCESSION	AE002301	AE002302	AE002160		
VERSION	AE002301.2 GI:8163203				
KEYWORDS					
SOURCE	Chlamydia muridarum.				
ORGANISM	Chlamydia muridarum				
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
AUTHORS	Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Ulterback,T., Berry,K., Bass,S., Linher,K., Weidman,D., Khouril,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., Deboy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39				
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)				
MEDLINE	20150255				
PUBMED	10684935				
REFERENCE	2 (bases 1 to 23773)				
AUTHORS	Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Ulterback,T., Berry,K., Bass,S., Linher,K., Weidman,D., Khouril,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., Deboy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Direct Submission				
TITLE	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
JOURNAL	On or before Jun 1, 2000 this sequence version replaced gi:7190382, gi:7190372.				

FEATURES

source	Location/Qualifiers
gene	1..23773
CDS	/organism="Chlamydia muridarum" /db_xref="taxon:83560" /note="synonym: Chlamydia trachomatis Mopn" 106..1842 /gene="TC0330" 106..1842 /gene="TC0330" 106..1842 /note="protein export protein, FliP family; identified by match to Pfam protein family HMM PF00771" /transl_start=1 /product="protein export protein, FliP family" /protein_id="AAF73547.1" /db_xref="GI:8163204"
translation	"MHIRKSGWKIVIPICILFTLVLPVLRWIIDFGVCANLACSLSIIIFVESIRSASARIPESTLLYCLRLGNLASTWIIASGVAAPLALGSFSLI

gene  
CDS  
/transl\_table=1  
/product="RNA polymerase sigma factor, sigma-70 family"  
/protein\_id="AAF39194.1"  
/db\_xref="GI:7190373"  
/translation="MKTODLADTQWLQVSTKEIHHRDFLEAYLPLVKNNAHRLASGM  
PSHVIEDLVASGVGLVRAVEREDPEKSRFESEYALFTIKAAIIDLRKODWPRSV  
YORANRLADAMDALROTIGKEPTDGLCEYINISOOELSHWSSRPALVSLINDDES  
CODEDEGLAERLADLERRAENGVDYIRKKEAISITLTERLLSLDEKEREQVMAIYYDDL  
VLKEGTGLGVSESRVSOIHSKALLKRLTSSL"  
2869. .4107  
/gene="TC0332"  
2869. .4107  
/gene="TC0332"  
/note="similar to SP:P00952; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="tyrosyl-tRNA synthetase"  
/protein\_id="AAF39195.1"  
/db\_xref="GI:7190374"  
/translation="MQQLIDSLQKRGILDNSSACLESITAPVASYLGDPAPAPLHIG  
HWIGICFLRLSAIGITPITALVGSATGIGPSSKRSVERSLDQEQVLDNSKILEVAL  
ANYLPDIRVNNADMGLSLMVDLRLDYGKFRIGSMKADVVKORYSEGISYTER  
SYLLDQSYDFALHFKHGVVLCQSGSDQMGNTSGIDYIRKGLGQAFLYPLTDS  
KGGIKGTSECTVWLDELPTSPYELFPLRISPOETPKIARMLTLDDDEVLADNR  
LENDQAVKRYVAIVIVKDVYAGAGLAAQALATTFESFANKCKNTSESRLAVOSGVC  
INVARADYIGKRMVDVYVQGFSGSSKGEARLIDQKGLYVQEPFLIDQSVLDGTILC  
FDRYLLSQGKKKQVIDLN"  
4117. .5556  
/gene="TC0333"  
4117. .5556  
/gene="TC0333"  
/note="similar to SP:P00349 PID:1193; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="6-phosphogluconate dehydrogenase,  
decarboxylating"  
/protein\_id="AAF39196.1"  
/db\_xref="GI:7190375"  
/translation="MAPADIGLGLAVGKNVLMITDGHFAVSVNRSEKTEEFK  
EHGENISLQGTALIEFVQSLAKRPRIIMIKAPAVDEMISLLPFLERDILLIDG  
NSYLYDSERRYIDLKKGILFVGMVSGEGEGARKPSIMPGNIEMAPYPIFOST  
AAVDGQPCSMWGTGGAGHVKVANGIEGDIOLICETMYEILKSRDLSIRDIGNI  
FEWNOTDLSYLSMAKSAVAVLTAKDENGVAASITIDVACOKGGRVAVADAKAGP  
MSLTIESVLARYLSAMKEVRQARREPPVASLITPQSEASVLIEDAREALYAKTIS  
YAOGEFLKQIISERNMDLNLGELALMRGCTIIQSAFLDKIHOGFESCPDAHSLMLO  
DYKRNVLINSETGFRRAIILHAVGAGVAPICLASALARYDGVRTFENSLPLVQGRDYF  
GAGYERODRPRGEFYHTDMLGSKNASRM"  
complement(5624. .7432)  
/gene="TC0334"  
complement(5624. .7432)  
/gene="TC0334"  
/note="similar to GB:D17650 SP:P37949 PID:1122398  
PID:436036 PID:1303804; identified by sequence similarity;  
putative"  
/codon\_start=1

gene  
CDS  
/transl\_table=1  
/product="GTP-binding protein Lepa"  
/protein\_id="AAF39197.1"  
/db\_xref="GI:7190376"  
/translation="MKPKIENINFSIAIHIDCKSTIADRLLESTIPEMRPCO  
LLDSMDLEREGITIKAPVMTYTEBSEYTELILITPGVDSYEVSNLSACBGA  
LLIIDAQGVQASLANVYLALEPDLLEITPLVNLIDIPAAQPEAIKQIEFFGLDTS  
NAIACSAKTGGQIPEILESIIRLVPKPPSQEFTLEKALIFDSHVDPVGMVVRVIS  
GEIKKDRITPMSTKSGSEFEVLGICAFLEPATILEGSLRAGOVGFYANLKKVYDKI  
GDVTYTKHPAKPELEPGEKELKPVFAGIYITDSDPDTLDALGRLOLNDALITIO  
ESSHSGFPGFCGLILHLEITFERISREPDITITAPSVIKYKVLKNGKTLFTN  
PTAYPDRALIHMEEPVHVYITTPQETLSNLSLCKDKRGICLKTMDLQHRVLVS  
ELPLNEIVSDNDKLSVTQVGSFDYRLGQVFGALIKLEILINDETVAESQVLR  
DKAASKRSICEKLVYDIPOLFKPIQOALNKKITIARETIRALAKVNTAKCYGDDIT  
RKRKLWKKQKKGRKMEKFCVSIPTAFVEVLMKE"  
complement(7617. .9206)  
/gene="TC0335"  
complement(7617. .9206)  
/gene="TC0335"  
/note="similar to GB:M28816 SP:P19568 PID:152470  
GB:AJ235269; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="Adp, ATP carrier protein"  
/protein\_id="AAF39198.1"  
/db\_xref="GI:7190377"  
/translation="MTQAEKPEFGLRSFLPMPIMHBLKKVLPFLMFCISFNYTL  
RDTKDLIVTAPSGSAEAIPIKMLVPSVAVMILYAKLSNITSQALFEYVLSF  
VPRFALFLVIYPRHITLHPDPAFDLQALITPSGFIETMLMRMPAAYVISEMG  
SVMLSPWGFANETITKISAKRYALFVGANVALLISCPAYVSSKRLASIGEVD  
PWGVTLYFLAMFSCAIIAACYWMNNRVLVDPNRYNAELKAKSRKMSGSEF  
SYLLSPYMLLALVLVICYGCVNLVYVTRKSQLQKPPNNEYSAFMGFSFMTGV  
SVFVMLFIGNVIRFGMLGALVTPMVAVTGAIFPALYIFDHATGLVAALGTTPL  
MLAVVGAONITISKSTKYALFEDATKEMAYPIIDQEQVKGKAIDVVAARFKSGS  
LIOOGLIVVGSISAMTPPLAVAFALIIWVLSATKLNLPLAASAKQDELAEATA  
AAEKESAPAKKEVSPALEGVS"  
complement(9533. .10029)  
/gene="TC0336"  
complement(9533. .10029)  
/gene="TC0336"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAF39199.1"  
/db\_xref="GI:7190378"  
/translation="MATROTTIQEIEIQLTKAISKYGSKENDCRVLPFGGIMH  
HFTLKMKHSAPEQLMLKTFVYIDSWTPRSINPKPRAPRGSKRRDFVNTFTDIER  
VLELARQVGDKDLARFSPKPLPSLKRILIRINNVSTELMNAVVALSQPTE"  
complement(10169. .10270)  
/gene="TC0337"  
complement(10169. .10270)  
/gene="TC0337"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAF39200.1"  
/db\_xref="GI:7190379"  
/translation="MOTNLNNLKHFIIFCPNOKISKMLRKKOKPL"  
10658. .11638  
/gene="TC0338"  
10658. .11638  
/gene="TC0338"  
/note="similar to PID:790546 PID:1777933 GB:AE000520;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="ABC transporter, periplasmic substrate-binding  
protein, putative"  
/protein\_id="AAF39201.1"

```

/db_xref="GI:7190380"
/translation="MFELHVRKRYHVLGELICLAGCEVLISSCGSGRGKSIDERLHIL
SMNRIYDCVRIKIDYKIVLIDGSDIPSYEMVDEDEDRM1ISQ1FPGNIGLIEH
SASLRKHEGNSKYLDIGARILDKNCFILSEDFPDHITMDGWM1SYXKEMASVL
VOQ1EYQAEEROKABQJLSMEDIDRMAVSLAT1PEKNRYLV7GHNAFSTFTRRL
SSDERESGNMKLRCKMSEPEGLSPBAQISIRIDIMVVEICANDGVFLEEDTLNODAL

Query Match      52.7%; Score 862.6; DB 1; Length 23773;
Best Local Similarity 72.5%; Pred. No.1.4e-195;
Matches 1147; Conservative 0; Mismatches 429; Indels 7; Gaps 2;

QY 6 AAAAACTATCAGATAGAAAATAAAGTATTTTCAGAGGT-AAATATGCAAAAACGA 64
Db 9252 ACAAGAAAACCTTACAAATAAACATAGCTTTTCAGAGGGTGAADATGATCAAAACGCC 9193
QY 65 AGAAAAACCTTTTGGAAAAATTTGGCTTTTCTTTGGCCGATACACTACTCAGAGCTAA 124
Db 9192 GGAAGAACCTTTTGGAAAAATTTGGCTTTTCTTTGGCCGATACACTACTCAGAGCTAA 9133
QY 125 GAAAGTTGCGCAATGTTCTTAATGTTCTTCTGATTAATTAATTAATCAATGAGTTCAG 184
Db 9132 GAAAGTTGCGCAATGTTCTTAATGTTCTTCTGATTAATTAATTAATCAATGAGTTCAG 9073
QY 185 CGATACAAAAGACACTCTTATTTGTGGAGCTCGTCTGTGGTGGAGGCAATACCTTT 244
Db 9072 AGATACAAAAGACACTCTTATGTCACAGACAGGCTGTGGAGGCAATACCTTTGAGT 9013
QY 245 CATCAAGTTTGGCTTTGTTGTCCTCGCTGCTATTTATCTTTATGCTTATTTATGCAAGCT 304
Db 9012 TATTAAGTTTGGCTTTGTTGTCCTCGCTGCTATTTATCTTTATGCTTATTTATGCAAGCT 8953
QY 305 AAGTATATTTTAAAGTAAGCAGGCTTATTTATGAGGAGGAGGAGGCTTTTATTTT 364
Db 8952 AAGCAATATTTTGAAGCAAGGCTCTTTCTATGACAGTCTCTCTCCATTTGCTTGT 8893
QY 365 CTTTGCCCTGTTCCGACGTATTTATTCAGCGATGTTTATACATCTTAAGT 424
Db 8892 TTTTCACAGTCTCTTTGTTGATTTACCATATGCTCATATTAATTAATTAATTAATTAAT 8833
QY 425 TGCTACCGCTTTACAGGCACTCTTACCTCAGAGATTTGCTAGAGCTGTTGCCATCTTAAG 484
Db 8832 CGCTATATACACTACAGGCACTCTTCCCTTGGTTCTTAAGCTTATCATGCTATGCTATG 8773
QY 485 AACTGACATTTGCGATTTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 544
Db 8772 CAACGTGACCTTTTGCAGCGCTTTTATGTTCTTCCGACCTTTTGGGAGAGTATGCTTTC 8713
QY 545 TCTAATGTTCTGGGATTTGCTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 604
Db 8712 GCTAATGTTCTGGGATTTGCAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 8653
QY 605 CGCTCTTTTGGTATCGGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTTGTTG 664
Db 8652 CGCTCTTTTGGTATCGGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTTGTTG 8593
QY 665 GGCTTCAAGTGAAGAGCTCGGTTTCTGAAGTGAAGTCTGTTGGGATTTCTTTAG 724
Db 8592 GTCTTCTAATTAATGCGGAGCTTGAAGTGAAGGCGAGACCCATG6GGAGTACTCTTTA 8533
QY 725 TCTTTGATGCTATGACTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 784
Db 8532 CTCTTAATGCTATGCTGCTGTTCTGTTGCAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8473
QY 785 CATATGAAGCTATGACCGATGCTGCTGCTTCTAATATCAGAGAAATGCAAAAGGGGAA 844
Db 8472 GATATGTTACGTTCTTACTGATCTAGATTTCTAATACCGAGAGCTGAAAA-----GC 8419
QY 845 AAAAGTGCTAAGCTAAAAATGAATGAATGAATGAGCTCCCTAATCTGATGATATC 904
Db 8418 TAAAAAATTTAAACCTAAGATGAGGCGAGAGCTTCTCTTATCTTTAAGATCCCC 8359
QY 905 TTATATCTTTTATTAAGCTCTGCTGTTATGCTATGCTATTTGCTATTAAGTATGCA 964

```

```

Db 8358 TTACATGCTTCTTACAGCTTCTGATTAATTTGCTACGAGTTTGCATCAACCTTGTGA 8299
QY 965 ACTGACTGGAAAGCTCAAGCTTAACCTGCAATATCTTAATATGATGACTATAGAGAT 1024
Db 8298 AATACTTGAAGAGCCACCTCAAAATGCAATTCCTTAATCCAAATGATATACAGGCAAT 8239
QY 1025 CATTGGGAACTCTCTCTCTGAGTGGGCTAGTATCCGTAATCAATCAATTTGTTGG 1084
Db 8238 CATGGGAACATTTCTCTTCTTGACAGGAGGTTTATCCGATTCGTTATGCTTATTG 8179
QY 1085 TGTAACTGATTCGTAAATTTGATGTTAACTGAGAGCCCTAGTCACTCTGTCATGAT 1144
Db 8178 TGTATGTTATTCGCTGCTTTGCTGTTTACAGAGGCTTATTAACCTGTCATGAT 8119
QY 1145 TCTCTAAGAGTATGCTTTTCTTGCTCTTTATCTTTAACAACCAAGCTTCTGGCT 1204
Db 8118 CCTGTAACAGAGAGGATCTTCTTGCTCTCTTATCTTCAAGATCATGCAACAGGAT 8059
QY 1205 GGTGCTATGTTGCGTACACTCTCTCATGCTAGCTGCTGTTGCGAGCTATACAGA 1264
Db 8058 GGTGCTAGCTTTGGAGACACTCTTATTAATGCTAGCGGTTGTTGGAGCCGTCAAA 7999
QY 1265 TATCTTTGAAATCCAAATACGCTCTCTTGACTCACTAATAAAGATGCTATAT 1324
Db 7998 TATCTTTCCAAATGACTAATATGCTCTTTGATGCACTAAGAGATGCTTACAT 7939
QY 1325 CCTCTTGACCAAGCAAAAGCTCAAGGTAGGCTGCTATGATAGTTGCCCGC 1384
Db 7938 CCTATTTGATCAAGAGCAAAAGTAAAGAAACCGCTATTAAGCTGTTGCTGCTAG 7879
QY 1385 CTTCGAAATACAGAGAGGCTTTAATCCAAAGGTTTGGTATCTGGAAGTAT 1444
Db 7878 ATTGCTAATATTTGGGAGATCTTTTATCCAAAGAGACTCTAGTTGTTGGGAAGCAT 7819
QY 1445 TGGAGCTATGACCCCTTATCTTGACAGTATCTTTTCACTATGCTATTTGGTGGT 1504
Db 7818 CAGTGCAGTATCCATTCCTGAGCGTAGCTCTTTGCAATTAATATGTTGGTTAAC 7759
QY 1505 TTCTGCACCTAAGTTTAACAACCTATTTAGCGAGTCTGCTTTAAGAACAAGAGT 1564
Db 7758 ATCTGCACCAACCTAATTAATTTGTTTACCTGCTTCAAGCTGCTAAGAACAAGAGCT 7699
QY 1565 GGTCAAGAAAGATTCAGCTCTG 1587
Db 7698 GGCAGAGGCCACAGCAGCTG 7676

RESULT 7
AX098423 1896 bp mRNA linear PAT 02-APR-2001
LOCUS AX098423
DEFINITION Sequence 2 from Patent WO0120009.
ACCESSION AX098423
VERSION AX098423.1 GI:13537715
KEYWORDS .
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 1896)
AUTHORS Reindl,A., Geigenberger,P.L., Neuhaus,H.E., Graeve-Kampfenkel,K.,
Moehlmann,T. and Tjaden,J.
TITILE Plants having altered amino acid contents and method for the
production thereof
JOURNAL Patent: WO 0120009-A 2 22-MAR-2001;
BASEF AKTINGESELISCHAFT (DE)
FEATURES
location/Qualifiers
source 1..1896
/organism="Solanum tuberosum"
/db_xref="taxon:4113"
BASE COUNT 473 a 369 c 445 g 609 t
ORIGIN

```

Query Match	23.2%	Score 380.2	DB 6	Length 1896
Best Local Similarity	55.1%	Pred. No. 3.1e-80		
Matches 804	Conservative 0	Mismatches 628	Indels 27	Gaps 2
OY	74	TTTTGAAATTTGGCCCTTTCTTCTGTGGCCGATACATACGACGATAAGAAAGTCT	133	
DB	285	TGTTGAAAAGAGACACCTAAGTTATATGGGATTTGAACTTGACCTTGAAGAAATAT	344	
OY	134	GCCATGTTCTTAATGTTCTTCTGATTTACATTAATACGAGTTCACGATACAA	193	
DB	345	ACCACCTGGGCGATGTTCTTTGATTCGTGTTAATATACATCCCTTAGGATACAA	404	
OY	194	AGACACTTATTTGGAGACCTCCGCTTCGTGTCAGACGATACCTTTCATCACT	253	
DB	405	GGATGTGTTGTTGATTAACAGCTAAAGGGTCCAGTGTGATTAATCCCTTCTGTAAC	464	
OY	254	TTGCGTTGTTGCCCTTGCTATTAATTTATGCTTATTAATGCAAGCTAAGTAAT	313	
DB	465	TTGGTGATTTGGCTATGAGCTATTTGATTCATGCTTTTGTACACAAAGTGGCTAAT	524	
OY	314	TTTAAGTAAGCAGGCTTATTTATGACAGTGGGAAGCCCTTTTAAATTTCTTGGCC	373	
DB	525	GTTTCAAGAGGAGCTTTTATACGTTATACCTCTTTTATGCAATTCCTTGGGGC	584	
OY	374	GTTCCGACTTAATTTATCCGCTACGAGATGTTTATATCCTACAAATTTGCTGACCG	433	
DB	585	GTTTGTGTTTGTGTTGATTAACCTTAAGCAATTTTACACCTTACAGCTTTGGTGAT	644	
OY	434	TTTACAGGCCATCTACCTCCAGAGTTGCTAGAGCTGTTGCCATCTTAAGAACTGAC	493	
DB	645	GCTTCTCAATACCTTGCTGCAAGATTTCTTGAGCAATTTCTTGTGAGAGATCTGAG	704	
OY	494	ATTTGCTGATTTTATGTTACTTGTGAATATGAGGGAAGCCATATCTTAATGTT	553	
DB	705	TTTCTGCTGTTGTTATGATGATGCTGAGCTTTGGGAAGTGTGGGTTAGTACTTT	764	
OY	554	CTGGGATTTCTTAATGAATTAACAAAATCCAGAGCAAGCTTTCTACGCTTTT	613	
DB	765	TTGGGGATTTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	824	
OY	614	CGATATGAGACTAATATTTCTTACTAGCTTGTGCTGCTCAATTTGTTGGGCTTCAA	673	
DB	825	TGGAATTTGAGCAATGTTGCTTATTTCTCTGTCGACAGTGAATCTTTTCTAG	884	
OY	674	GTTGAGAGCTTCGTTTCTGAAGTGTAGATCTTTGGGGAATTTCTTACGCTTTT	733	
DB	885	CTTGAGAGGCTTTAGGCTCGAGATGATGTTGGGCTATCCCTGAAAGAAATGAT	944	
OY	734	GGCTATGACTATGATCTGAGCTTGTCTTATGCGCAGTACTGCTGATCAATAGAA	793	
DB	945	GAGTATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1004	
OY	794	CGTATGACGATCTGCTTCTTATATCCAGAAATGCAAAAGGAAAAAGGTGC	853	
DB	1005	TGTTGCTCTCCCAATCTGTAGCA-----GAAAGAAAGGT	1040	
OY	854	TAAACCTAAATGAATGAAGATAGCTTCTTATCTTGAATAGATCTCTTATATCT	913	
DB	1041	AAAACCTAATGACACCAATGAGAGCTTGAAGCTTGTGCTCTTCAAAATATATCAG	1100	
OY	914	TTTATTAACCTCTTGTGTTATGCTTATGATTTGATTTGATTAATCAATGACCTTG	973	
DB	1101	GGATCTTGACACATTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT	1160	
OY	974	GAAAGCTGACCTGAATGCAATCTTAATATGATGATGATGATGATGATGATGATGAT	1033	
DB	1161	GAACTCAAACTCAAACTCAATGTTCCCAAGCCCAATGATATCTTCAATTCATGGTGA	1220	
OY	1034	CTTCTCTCTTGAGCTGGCGTAGTATCCGATATCAATGATATTTCTGATGATGAT	1093	
DB	1221	CTTCTCAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1277	
OY	1094	CATTCGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1153	

DB	1278	TTTCGACAGATATGATGGGAGAGCAGCAGCAATACACTACAGCTTGCTCTTAC	1337	
OY	1154	AGTATCTTTTCTTCTGCTCTTTGATTTAGAAACCAAGCTTCTGGGCTGGCTAT	1213	
DB	1338	CGAGTTGTTCTTCTCTCTGCTTTGTTGGGCGACCTCTAGACCTTCTTGGCAA	1397	
OY	1214	GTTGCTGACACTCTCTCATAGCTAGCTGTGTTGTCGAGCTATACAAATATCTTTC	1273	
DB	1398	GTTTGGATGATCTCTCTTCTAGACGCTGCTATGTTGGTGCATGCAATTTTCAG	1457	
OY	1274	GAAATCCAAATATGCTCTTGTACTCAACTAAGAAATGAGCTATATCCTTTGA	1333	
DB	1458	TAAAGTGCAGAAATATGTTGTTGACCCCTGCAAGAAATGAGCTATCTTTGGA	1517	
OY	1334	CCAAGGCAAAAGCTAAGGCTGATGATGATGATGATGATGATGATGATGATGAT	1393	
DB	1518	TGAGGACACCAAGTTAAAGGAAAGGAGCAATGATGTTGTGCAATCCATGGGAAA	1577	
OY	1394	ATCAGAGAGACTTTAATCCAAAGGTTTCTGTTATCTGTGAAGTATTTGAGCTAT	1453	
DB	1578	GTCGTGAGAGACTTTGATACAAAGTTGATGATTTTGTGCTTGTGCTGACGCTC	1637	
OY	1454	GACCCCTTATCTTGACGATGATTTCTTTTATCATTTGCTATTTGGTTGTTGCAAC	1513	
DB	1638	GACACCTTACCTTGGCGTGTCTTATGATTTGTTCTTGCATAGTTGGTGGAGCAGCAA	1697	
OY	1514	TAACTTAACAAACTATTG	1532	
DB	1698	GTCCTTGGATGAGACATTC	1716	

RESULT 8  
STPLATDR 2300 bp mRNA linear PLN 22-JAN-1999  
LOCUS Solanum tuberosum mRNA for plastidic ATP/ADP-transporter.  
DEFINITION Y10821 GI:4138582  
ACCESSION Y10821.1  
VERSION plastidic ATP/ADP-transporter.  
KEYWORDS potato.  
SOURCE Solanum tuberosum  
ORGANISM

REFERENCE  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 2300)  
Tjaden,J., Mohlman,T., Kampfenkel,K., Heinrichs,G. and Neuhaus,H.E. Altered plastidic ATP/ADP-transporter activity influences potato (Solanum tuberosum L.) tuber morphology, yield and composition of tuber starch  
Plant J. 16, 531-540 (1998)  
2 (bases 1 to 2300)  
Neuhaus,E.  
AUTHORS

JOURNAL  
REFERENCE  
TITLE Submitted (24-JAN-1997) E. Neuhaus, University of Osnabrueck, Plant Physiology, Barbarastrasse 11, D-49069 Osnabrueck, FRG  
REMARK  
3 (bases 1 to 2300)  
Revised by [3]  
JOURNAL  
REFERENCE  
TITLE Submitted (06-JAN-1999) E. Neuhaus, University of Osnabrueck, Plant Physiology, Barbarastrasse 11, D-49069 Osnabrueck, FRG  
COMMENT  
On Jan 9, 1999 this sequence version replaced gi:2326873.  
FEATURES  
location/Qualifiers  
1..2300  
/organism="Solanum tuberosum"  
/cultivar="Desire"  
/db\_xref="taxon:4113"  
127..2022  
/codon\_start=1  
/product="plastidic ATP/ADP-transporter"  
/protein\_id="CA71785.1"  
/db\_xref="GI:4138583"  
/db\_xref="SPTREMBL:024381"

CDS

```

/translation="MEGVLOTFRGLSLSPKIKAFPLPQGLNRENSLSLKNP
LNGVLSNSGFOKVOGPDIPOLFGOKRCEPCIEAABAAAAGADGOLFEXDOPK
FMGIELATLKRIIPGAMFECILFNTLLRDKDLVLYPLSLFHEPFAADLNL
MAIGFMLYTKLANLSKALPYVILPIAFGAEVLYLPLSLFHEPFAADLNL
TLGPRFLBIALIRIWSFCLFYMAELMWSVSVLFWEFAOITVLEAKFEPYLFJ
LGNALJESGRTVYFSLRSLSLPGVGNALISLKGMSIYVMGGACAFEPYVNR
NVALPTRSKKKKPMNTMISLAKFLVSKYLRDIALVLAAGIINIVEVWKSLK
AOFSPNESSFMGPDSTATGIATFTMMLSQWIDCTGAWGAARITPVLITGVGF
PSLLFGAPLAPLAKFGMTPLLAIVYCAAMONITSKSKYSLPDPCKEMAITPDED
TKVCKAIDIVYCNPLGSGALLIOFPLITFSLASSPVLGVLVIVLWMLDAK
SLDGQFOTLROEDLEKEMERASLIKIPVVSOMENGNPLSSESLNPAGDSTMASSE
PSSPSRL"
BASE COUNT      590 a      443 c      518 g      749 t
ORIGIN
Query Match      23.2%; Score 380.2; DB 8; Length 2300;
Best Local Similarity 55.1%; Pred. No. 3.1e-80;
Matches 804; Conservative 0; Mismatches 628; Indels 27; Gaps 2;
OY 74 TTTTGAATAATGCGCTTTCTTGTGGCGGATACATACAGCAGCTAAGAAAGTCG 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 411 TGTGAAAAGAGACAACTAAGTTTATGGGATTTGACCTTAAGAAAATTAT 470
OY 134 GCCAATGTTCCATATGTTCTTGTATTAATTAATACGCTGTACGCGATACAA 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 471 ACCACTTGGGGGATGTTCTTTGTATTCGTTTAATTATCAATCCTTAGGATAC 530
OY 194 AGACACTCTTATTTGGAGCTCTGCTTGTGTCAGAGGCAATACCTTTCATCAAG 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 531 GGATGTGTGTGTGTACAGCAAAAGGTCACAGTGTGATTAATCCCTTCTTGA 590
OY 254 TTGGCTTTGTGTCCTGCTATTAATCTTAATTAATTAATTAATTAATTAATTA 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 591 TTGGGTAATTTGCCATATGCTATGATTAATGATTAATGATTAATGATTAATG 650
OY 314 TTTAAGTAAGCAGCCTTATTTATGTCAGTGGAGACGCCCTTTTATTTATTTT 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 651 GTTGCAAGAGAGCTCTTTTATATCTGTATTAATCTTCTTATTAATGATTT 710
OY 374 GTTCCGAGCTATTAATTAATCCGTACGCGATTTTATACCTTAAGAAATTTG 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 711 GTTGGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 770
OY 434 TTTAGAGGCATCTACCTACGATTCGATGAGAGCTGTCATCTTAAGAAAGC 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 771 GCTTCAATATCCCTTGCTGCAAGATTCTTGGAACCAATGCTATTCGAGAT 830
OY 494 ATTGCTGATTTTATGCTACTGCTGCAATATGGGAGAGCGCTATCTTAATG 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 831 TTTCTGCTTGTCTATGTCATGCTGAGCTTTGGGAGAGTGTGTTCACTACT 890
OY 554 CTGGGATTTGCTAATGAATAATTAACAAGCAAGCAAGCAAGCTTTTACCTCT 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 891 TTGGGATTTGCTAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 950
OY 614 CGGTATCGAGCTAATATTTCTTACTAGCTGTGCGCAAGTAATGTTGGCTTCA 673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 951 TGGACTTGGAGCAATGTTGCTTATTTTCTGTGTCGACAGCTAAGTACTTTCT 1010
OY 674 GTTGAAGCTTCCGTTTCTGAAGCTAGATCTTGGGGAATTTCTTAAGCTTT 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1011 CTGAGAAGCTTTTATGAGCTGAGTGTATGTTGGGCAATGCTCCCTGAAGA 1070
OY 734 GGTATGACTATTTATGAGCTTGTCTTAATGCGAGTACTGCTGATCAATAGAA 793
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1071 GAGTATGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATA 1130
OY 794 CGATATGACGATCTCGCTTATATATCAAGAAAGCAAGAAAGGAAAAAGCTG 853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1131 TGTGCTCTCCCACTGAGCA-----GAGAGAGAGG 1166
OY 854 TAAACCTAAATGAATATGAAGATGCTCTCTATCTTGAATAGATCTCTATAT 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 1167 AAAACCTAACATGACCAACATGAGAGCTTGAGATTCTGTGCTCTTCAAAATATAT 1226
OY 914 TTTATTAATCTCTTGTTGTTATTTGCTTAATGTTGATTAATTAATGAAGTCTG 973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1227 GATCTTGGCACATTTGTTGATCATATGATGATTAATCAACCTGTTGAACTTACATG 1286
OY 974 GAAAGCTGCTAAAGCTGCAATATPCTTAATATGATGATTAATGAGTCTGAGTGGG 1033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1287 GAAGTAAAGCTAAAGCTCAGTTCCCAAGCCCAATGATCTCTCATTTAGGGTGA 1346
OY 1034 CTTCCTCTCTGAGCTGGCGTAGTATTCGTAATCATATGCTATTTGTTGGTGAAGCT 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1347 CTTCCTCACTGCTAGGAAATATGACACTTTACAAATGATGTTGTTAAGCAATAGA---T 1403
OY 1094 CATTCGTAATTTGATGATGTTAATGAGCCCTAGTCACTCTCTGATGTTCTCTAAC 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1404 TTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1463
OY 1154 AGGTATGCTTTCTGCTCTGCTTTGATCTTTAAGAAACCAAGCTTGGGCTGTCGTAT 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1464 CGAGTGTGTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1523
OY 1214 GTTGGTACAACTCTCTCATGCTAGCTGTGTTGCGAGCAATACAGAAATATCTTTC 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1524 GTTGGAAATGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1583
OY 1274 GAAATCACAATATACCTCTCTTTGATCTCACTAAAGAAATGCGCTATATCCCTCTGA 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1584 TAAGATGCAAAAGATATATATATATATATATATATATATATATATATATATAT 1643
OY 1334 CCAGAGCAAAAGATGAAAGCTGATATGATGATGATGATGATGATGATGATGATG 1393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1644 TGAGGACACCAAGCTTAAGGAGAGGAGCAAGATGATGATGATGATGATGATGATG 1703
OY 1394 ATCAGAGGAGCTTTATATCAACAAAGTGTGCTGTTATCTGTGAAGATATGAGCTAT 1453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1704 GTCTGAGGAGCTTTGATACAAAGCTGATGATGATGATGATGATGATGATGATG 1763
OY 1454 GACCCCTTATCTTGCAGTATTTCTTTTCAATCATGCTATTTGTTGTTCTGCAAC 1513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1764 GACACCTTACCTTGGGCGGTGCTCTAGTAATGTTGCTGCAATGTTGGAGACCA 1823
OY 1514 TAAGTAAACAACTATTC 1532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1824 GTCTTTGATGACAGTTTC 1842

```

RESULT 9  
 GSU251356  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATUES  
 source

GSU251356 2126 bp mRNA linear PIN 04-JAN-2002  
 Galdieria sulphuraria chloroplast mRNA for plastidic ATP/ADP  
 transporter (atp1 gene).  
 AJ251356  
 AJ251356.1 GI:18073247  
 atp1 gene; plastidic ATP/ADP transporter.  
 Galdieria sulphuraria.  
 Placid Galdieria sulphuraria  
 Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;  
 Porphyridiaceae; Galdieria.  
 1  
 Stamme, C. and Neuhaus, E. H.  
 Analysis of the plastidic ATP/ADP transporter from the red algae  
 Galdieria sulphuraria  
 Unpublished  
 Neuhaus, E. H.  
 2 (bases 1 to 2126)  
 Direct Submission  
 Submitted (26-NOV-1999) Neuhaus E. H., Plant Physiology, University  
 of Osnabrueck, Barbarastr. 11, 49069 Osnabrueck, GERMANY  
 Location/Qualifiers  
 1..2126  
 /organism="Galdieria sulphuraria"  
 /organelle="plastid"

```

gene
CDS
1. 2018
/gene="atp1"
<1. .2018
/gene="atp1"
/codon_start=3
/transl_table=1
/product="plasticidc ATP/Adp transporter"
/protein_id="CAC80882.1"
/db_xref="GI:18073248"
/translacion="EMVQPSSEFVNLVHSSHMKGTKRQSVSRKNNAVHFCPKETSRF
SQAGRRCCKHKGSEETFYRNEAASNNELSTAAVEISEHERIALSSYPVPPQIH
SNOSTRSLATATVATIGQWFKLVKSSKADVDLLSPQFLQNASGTSVSGSVAG
DGIKALPKRNEIKKKLPIGWFPMILFENYTLIDPKYVILVTTGGATIIPELTYNIA
LPGVLTIAASKLSNMDRETLFYVCVPIELPIISAFLYELRLAHLPYAVDNTI
AGYIPASFEAPLGIIRNNTKRLFTYTLAEKMSVYSLLEWFAVEVSVDAKATYPI
FGLVANVALITSGQVYRFVSDIRKKLPBGVDGWSLKYLTMTIGVAGSVTILMLVYI
OKNVLTPDNCVDMSKRKNTKQKRLGKESAVYLAQSVIRNLALVYIAYGMSINIVEE
VSNMKSLEAFPPDPSNSTEWMGFSTGTSVTLIMMLGRIIPKFGGFALATTPVTVY
LGITGLFEELCTLFSROLHPVVMGIGTPTMLAVLGAAGNAIISKSSKSYLDFDCKEK
AYITPLDAEOKSKGAATIDVIGPMKSGGSGFIQOGLIPAGVSLAASPXYLAFILFYI
GMATLAARSLDKQFOEAMKRREVAEVAWAKTIQDPTEAAKVAKESKQ"
BASE COUNT
567 a 388 c 467 g 704 t
ORIGIN

```

Query Match	22.9%;	Score 374.8;	DB 8;	Length 2126;
Best Local Similarity	55.6%;	Pred. NO. 6e-79;		
Matches 790; Conservative	0;	Mismatches 617;	Indels 15;	Gaps 3

OY	115	ACGAGCAAAAGGTTCTCGCAATGTTCCGTAATGGTCTTCTATTAACCTTAACCTA	174
Db	514	ACGAGCTGAAAAAGATCTCCTCGTTGGGTGCATGTTTTATGATTTCTATTCATTTA	573
OY	175	CGGTGTTACGGATACAAAAGACATCTTAATGTGGAGCTCCGTGGTTCGTGCAGAG	234
Db	574	CTATCTCGAGAGATACAAAAGACCTCTCTGACAGCAACTACAGGA-----CGAGAA	627
OY	235	CAATACCTTCATCAAGTTTTGGCTGTGTGCCCTGTGTAATCTTATCTTATCTATT	294
Db	628	TCATACCGTTCCTTAATAACATATGCAAATCTCTCGAGAGCTTAATTTACTATTCGCT	687
OY	295	ATGCAGAGCTAAGTAATATTTTAAGTAAGCAGGCTTAATTTATGCAGAGGGAACGCTT	354
Db	688	ATTCGAAGTTATCAATATGTTTGATCGGAAACCTGATTTTAATGTTGTATCGTTCCCT	747
OY	355	TTTTAATTTTCTTGCCCTGTGCCGACGTAAATTATCCGCTACGCCGATGTTTTACAT	414
Db	748	TTTATATTATTTTTCATATCTTGCTTTGTGTATACCCGTTGAAGACATGCATCTTATC	807
OY	415	CTACAGATTTTGCGACCGTTTACAGGCCATCCACTCCAGATTTGCTAGACTGTG	474
Db	808	CTTATGCCCTTGTGGATTTGGATAGCTGGTATCTTCCCGTCTCTTTTGGCTCCCTTGG	867
OY	475	CCATCTTAAAGAACTGACATTTTGCTGCATTTTATGTACTTGTGCTGACATAGGGAAAGC	534
Db	868	GAATCATTTCCGAATTTGGACTTTTGCTTTCTTACACCCCTGGCGAATATAGGGATCTG	927
OY	535	TCATGCTATCTAATTTTGTGGGGAATTTCTATGAATAATTAACAAAATCCAGAACGA	594
Db	928	TTGTGTTATCTCTGCTGTTTTGGGGTTTTTGCTATATGAAGTAACCTCAGTAGATAGGCA	987
OY	595	AGCCTTTCTACGCTCTTTTGGTATCGGAGCTAATATTTCTTACTAGGCTGTGTCGCG	654
Db	988	AAAGATATTATCCCTATTTTGGCTTGGTTGCGAATGTTCGTATTTTTTTCAGGGCAAT	104
OY	655	CAATGTTTGGGCTCAAAAGTTAGAGACTTCGTTTCTGAAGGTGATATCCTTGGGGAA	714
Db	1048	ATGATACATTTGTTTCGATATTTCGTCGAAGTTACACCTGGCTGTGATGTGGGGCTG	1107
OY	715	TTTCTTACGCTTTTGATGGGTATAGCTATGTATCTGAGCTGTGTTTATAGGCCAGTT	774
Db	1108	TTTTCATTGAATATTGTGATACATGTAAGAGATCTGCTGGTTCAGTCATCATTTCTTGCA	1167

QY	775	ACTGTGATCATTAATGAAGACGTATTGACCGATCTCCGCTCTATTAATCCAGAAAGAAATGC	834
Db	1168	TGCTATATATTCAGAAAATATGCGACCGATCC-----CAATTGCTGTGAATGAGTA	1221
QY	835	AAAAGGCGAAAAAGGTGCTTAACCTAAATGAATGAAATGAAATAGCTTCCTATCTTG	894
Db	1222	AAAGAAACACCTTAACCAAAAGACCAGCGTGGCTTAAGAGAGATGCTGTTATCTGG	1281
QY	895	ATGATGCTTCCTATATTCCTTTATTTAACTCTCTGGTATTTGCGTATGATTTGGCATTA	954
Db	1282	CCAACTGACCTTAATATTCGGAACCTTGCGCTTTGTGTATGCTATGATGATGTATGATTA	1341
QY	995	ACTTAATGAGTGTGCTTGGAAAAGTCAGCTGAACCTGCAATATCTTAATGATGATGACT	1014
Db	1342	ATATTTTGGAAATCATCATGAGGACCAAGTGAAGAAAGAACCTTTCTGATTCCTAATCTT	1401
QY	1015	ATAGGAGTTCATGAGGGAACCTCTCTCTGAGCTAGGCGCTAGTATCCGACTTATCAATGC	1074
Db	1402	ATTCAACCTTTATGSGTGTGCTTCTCACTGTACTGCTGTTGCGTAAAC--ACTTATCATTA	1458
QY	1075	TATTTGTTGGTGTGTACGTCATTTGCTAAATTTGGATGTTAACTGAGACCCTAGTCACTC	1134
Db	1459	TGCTTCTTGTCGATTTATTTTCCAAAAGTTGGTTGGGAGTTTGCAGCTCTTATTACTC	1518
QY	1135	CTGTGATGCTTCTCTTAACAGGTATGCTTTCTGTGCTCTGTATATCTTTAAGAACCAAG	1194
Db	1519	CAACTGTTTGGGATTAATCTGGCTTTTATTTCTGTTTAACTCTCTTAATCTGCGCAAT	1578
QY	1195	CTTCTGGGCTGATGCGCTATGTTGCGGTACAACTCTCATGCTAGCTGAGTGTGTCGGAAG	1254
Db	1579	TGCATCCAGTGTGGATGCTGTGAACGACACCTTTGATGCTTGCAAGTCTGTGTTGGTG	1638
QY	1255	CTATACAGATATTTCTTTGCAAAATCCACAAAATPACGCTCTCTTGACTCAACTAAGAAA	1314
Db	1639	CTGCACAAAATATTTCTTCTTAATGTGACACAAATATTGCTGTGCAATCCTTGCAAGAAA	1698
QY	1315	TGGGCTATATCCCTTTACCAAGGCAAAAAAGTCAAAAGTAAAGCTGCTATATGATGAG	1374
Db	1699	TGGCTTATATTTCCATTTGAGCCCTGGAACGAAATCTTAAGGAAAGGCTGCAATCGATTA	1758
QY	1375	TTGCGCGCCGCTTGGGAAATACAGAGAGAGCTTTATATCAACCAAGGTTTGCTCGATTACT	1434
Db	1759	TTTGGCAATCCATGAGGCAAGTCTGTGTGCTGCTTATTTCAACAGGCTCTATTTTTCGAG	1818
QY	1435	GTGAGAGTATTGAGAGCTATGACCCCTTATCTTTCGACGTATCTCTTTTCATCATTTGCTA	1494
Db	1819	TGGGTTCTTTAGCAGCCCTCAACACCCTATCTTGCTTTATATGTTGTTATTTGTTGGTA	1878
QY	1495	TTTGGTGGTTCCTGCAACTAAGTTTAAACAAACCTATCTTGAAG	1536
Db	1879	TGTGGATGACAGAGCTCTGTTCTTTAGCAACCAAGTTTCAAG	1920

RESULT 10	
AY098893	
LOCUS	2122 bp mRNA linear PLN 21-MAY-2002
DEFINITION	Citrus hybrid cultivar plastidic ATP/ADP transporter mRNA, partial cds; nuclear gene for plastid product.
ACCESSION	AY098893
VERSION	AY098893.1 GI:21063924
KEYWORDS	
SOURCE	Citrus hybrid cultivar.
ORGANISM	Plastid Citrus hybrid cultivar
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurossids II; Sapindales; Rutaceae; Citrus.
AUTHORS	1 (bases 1 to 2122)
TITLE	Li,C., Weiss,D. and Goldschmidt,E.E.
JOURNAL	Citrus mRNA for plastidic ATP/ADP transporter
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 2122)
TITLE	Li,C., Weiss,D. and Goldschmidt,E.E.
REFERENCE	Direct Submission









D <b>b</b>	1386	TTTGCTATTACAGTGGTGTGGGTTCTTCTCTCTATATATTTGTTGGCGGCCCATTTGGCAC	1445
O <b>y</b>	1201	GCGTGGCTATATGTTGGGATACAACTCCTCTCATCTAGCTGTGGTGTGGAGGCTATAC	1260
D <b>b</b>	1446	CACATTGTGCCAAGCTTGATGTATGACACCGCTACTTGACAGTGTGATGGTGGCCCTTC	1505
O <b>y</b>	1261	AGATATTCTTTCGAAATCCACAAATATCGCTCTCTTGACTCAACTAAGCAATAGCCT	1320
D <b>b</b>	1506	AGAAATATCTTCAGCAAGAGTGGCAAGTACAGCTTTTGACCCCTTGCAAAAGAAATAGGCT	1565
O <b>y</b>	1321	ATATCCCTCTTGACCAAGGCAAAAAGTCAAAAGTAAAGGCTCTATGATAGTGGCG	1380
D <b>b</b>	1566	ATATCCCATTTGATGATAGGACACCAAGGTTTAAAGGCAAAAGCTGCGATGACGTTGCTGCA	1625
O <b>y</b>	1381	CCCGCTTCGGAAATACAGAGAGGCTTAAATCAACAAGTTCCCTGCTTATCTGTGAA	1440
D <b>b</b>	1626	ACCCATTATGGGAATACAGGGGAGGCTTTAATACAGACGTGATGATCTTATCCCTTGGAT	1685
O <b>y</b>	1441	GTAATTGAGACTATGACCCCTTATCTTGCAAGTATCTTCTTTTCAATATGCTATTGGT	1500
D <b>b</b>	1686	CACATGCAATTCACAGCGCGTATCTAGGAATGATCTTGTGGTTATGTGCATCGGTGGT	1745
O <b>y</b>	1501	TGGTTTCGCAACTAAGTTAAACAACATTCAT	1532
D <b>b</b>	1746	TAGCTCAGCTTAGTGCCTGGAGGACAGTTTC	1777

RESULT 12	
ATANTMR	2181 bp mRNA linear PLN 26-NOV-1999
LOCUS	
DEFINITION	A.thaliana mRNA for adenine nucleotide translocase.
ACCESSION	Z49227
VERSION	249227.2 GI:6469339
KEYWORDS	adenine nucleotide translocase.
SOURCE	Arabidopsis thaliana.
ORGANISM	Arabidopsis thaliana.
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE	1 (bases 1 to 2181)
AUTHORS	Kampfenkel,K., Mohlmann,T., Batz,O., Van Montagu,M., Inze,D. and Neuhaus,H.E.
TITLE	Molecular characterization of an Arabidopsis thaliana cDNA encoding a novel putative adenylyate translocator of higher plants
JOURNAL	FEMS Lett. 374 (3), 351-355 (1995)
MEDLINE	96069943
PUBMED	7589569
REMARK	(sites)
REFERENCE	2 (bases 1 to 2181)
AUTHORS	Kampfenkel,K.K.
TITLE	Direct Submission
JOURNAL	Submitted (05-MAY-1995) Kampfenkel K.K., Universiteit Gent, Laboratorium voor Genetika, K.L. Ledeganckstraat 35, Gent, Belgium, B-9000 Gent
REMARK	revised by [3]
REFERENCE	3 (bases 1 to 2181)
AUTHORS	Kampfenkel,K.K.
TITLE	Direct Submission
JOURNAL	Submitted (26-NOV-1999) Kampfenkel K.K., Universiteit Gent, Laboratorium voor Genetika, K.L. Ledeganckstraat 35, Gent, Belgium, B-9000 Gent
COMMENT	On Nov 27, 1999 this sequence version replaced gi:1051108.
FEATURES	location/Qualifiers
SOURCE	1..2181
	/organism="Arabidopsis thaliana"
	/strain="Var. Columbia"
	/db_xref="taxon:3702"
	/clone="C23"
	/rissue_type="Whole seedling"
	/clone_lib="lamdays"
	/dev_stage="seedling"
	110..1981
	/codon_start=1
CDS	

[illegible]

Oy	901	CNCCTATTTTCCTTATTAAACCTCCTCGGTATTTGCCATAGTATTGGCATTAACCTAA	960
Db	1200	CACCATACATTTAGAGATCTTGCTACTTTTAGTGGTGCATACGGTATTGATCAACAATCTTG	1259
Oy	961	TGGAAGTAGCTTGGAAAACCTCAGCTGAACCTGCATAATCTCATATAGAATGACTATAGTG	1020
Db	1260	TGGAAGTACATGGAANAATCAAAAGCTTAAAGCTCAGTTCCCTAGCCCCGAATGAGTACTCAG	1319
Oy	1021	AGTTCATGGGGAACTTCTCTCTTGAGCTGGCGTAgTATTCGACTTATTCATGCTATTG	1080
Db	1320	CATTTATGGGAACACTTCTCAACCTGCACGGGCTGTGCAACATTCACAAAGATGC---TTC	1376
Oy	1081	TTGGTGGTAAAGCTCATTCGTAAATTGGATGGTTAATCGAGCCCTAGCACMCCCTGCA	1140
Db	1377	TCAGCCAACTAGCTATTTCATTAAGATGTTGGGGAGTAcCTGCAAAAGATCACCCCACATG	1436
Oy	1141	TGGTTCTCCTAACAGGTATCTTTCTTCGCTCTGTATTCCTTTAGMAACAAGCTTCG	1200
Db	1437	TTCTCTATTAGACTGGTGGTCTTCTCTCTAATATTGTTTGGCGCCCATTCGCGAC	1496
Oy	1201	GGCCTGGCTATGTTCCGTACAACCTCCTCTATGCTAGCTAGCTGGTGGTGGAGCTATAC	1260
Db	1497	CAC TTGTTGCCAAGCTGGTATGACACCGCTACTTCACACTGTGATATGTCGGTGCCCTTC	1556
Oy	1261	AGAATATTCTTCGAAATTCACAANAATAGCTCTCTTCATCAGCACTAAGAANAATGGCT	1320
Db	1557	AGAATATCTTCAGCAAGAGTGCACAGTACAGCTTGTTCGACCCTTCGAAAAGAAATGGCT	1616
Oy	1321	ATATCCCTCTTGACCAAGAGCAAAAAAGTCAAAAGTAAAGGCTCTATGATGATGTCGG	1380
Db	1617	ATATCCCATTTGGATAGGACACCAAGAGTTAAAGGCAANACCTCGATGTGACGTGCTGCA	1676
Oy	1381	CCCCCTTGGAAATACAGGAGAGCGTTTAATCCACAANAAGTATTCGCTTACTCTGGGA	1440
Db	1677	ACCCA TTAAGGGAATCAGGGGAGCTTTAATCAGCAGCTTCATGATCTTATCTTTGGAT	1736
Oy	1441	GTATTGAGACTATGACCCCTTATCTTCGAGTATCTCTTTTCAATCATGCTATTGCT	1500
Db	1737	CAC TAGCAATTCACAGCGCGCTATCTAAGGAATGATCTGTTGGTTATTGTCACTGGGTGT	1796
Oy	1501	TGGTTTCTGCAACTAAGTTAAACAACACTTTC	1532
Db	1797	TAGCTGACGCTAAGTGCCTGAGGAGCAAGTTTC	1828

LOCUS	AY045903	2151 bp	mRNA	linear	PLN 24-APR-2002
DEFINITION	Arabidopsis thaliana putative adenine nucleotide translocase (At1g80300) mRNA, complete cds.				
ACCESSION	AY045903				
VERSION	AY045903.1	GI:15028092			
KEYWORDS	PLN_CDNA				
ORGANISM	Arabidopsis thaliana.				
SOURCE	Arabidopsis thaliana				
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae: eucotsids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 2151)				
AUTHORS	Yamada, K., Iiu, S.X., Sakano, H., Pham, P. K., Banh, J., Chung, M. K., Goldsmith, A. D., Lee, J. M., Quach, H. L., Toriumi, M., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T., Saou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R. W., Ecker, J. R. and Theologis, A.				
TITLE	Arabidopsis Full Length cDNA Clones				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2151)				
AUTHORS	Yamada, K., Iiu, S.X., Pham, P. K., Banh, J., Dale, J. M., Goldsmith, A. D., Jiang, P. X., Lee, J. M., Onodera, C. S., Quach, H. L., Tang, C. C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P.,				

TITLE		Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koeseema, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S. E., Shinozaki, K., Davis, R. W., Ecker, J. R. and Theologis, A.	
JOURNAL		Submitted (10-Jul-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA	
COMMENT		RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RafL cDNAs (RafL cDNA : RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.	
<p>The Sak, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RafL cDNAs: Yamada, K., Liu, S. X., Pham, P. K., Banh, J., Dale, J. M., Goldsmith, A. D., Jiang, P. X., Lee, J. M., Onodera, C. S., Quach, H. L., Tang, C. C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koeseema, E., Lam, B., Lin, J., Meyers, M. C., Miranda, M., Nguyen, M., Palm, C. J., Shinn, P., Southwick, A., Tracy, S. E., Davis, R. W., Ecker, J. R. and Theologis, A.</p> <p>Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.</p>			
FEATURES		Location/Qualifiers	
source		1. .2151	
		/organism="Arabidopsis thaliana"	
		/db_xref="taxon:3702"	
		/chromosome="1"	
		/note="This clone is in a modified pbluescript vector (pLC-1) as a BamHI/XhoI insert.	
gene		1. .2151	
		/gene="Atlg80300"	
5'UTR		1. .27	
		/gene="Atlg80300"	
CDS		28. .1902	
		/gene="Atlg80300"	
		/codon_start=1	
		/evidence=experimental	
		/product="putative adenine nucleotide translocase"	
		/protein_id="AKK76577.1"	
		/db_xref="GI:15028093"	
		/translation="MEAVITQFGLTSPTKPIGVRSQLOPSHGKLRLEPAKPRMLHQLSLNSGHRKRPQTEPTLHGISISHRKESTETICAEAAADGAVFEGDSAAVAAVAAPIKEVEVALTKIILPLGMEFCILFNTITDRDTDAVGVNPPSSAEIIPILKTMVAPLMAIGFLITLTKLSNLSKALFVTYIVPTITGACGVFAPKSNITHPALADKLTTGAPRMGPAILIRISFCLEFYMAELGMSVVSVLPWGFANDITVDEAKKTPYFGLGDNALIFSGRTVKYFSLNRKNGGVDGMAVSLAAMSIVVGMGLAICLAWMSNRYVLPFLRSNKKRKPMPGTMESLKLPLVSSPYIRDLATLVAVGISINLVEYTWKSSLKAQPSBNEYSAEMGDSTCTGATVPTFMILSOYVFNKYGVAKATTPPYLVLTGSAFSSILFEGFAPLPVARTGLMPLAAVYVGLGVNIFPSKAKSYLDFDCKEMAYPLPLAEDTYKKKAAIDVYCNPLKSGGALIQPMILRSGLANSPIYLMLITLVITANLAAAKSLDGQPNLSRSEELKEKEMRASVAKIPVAVSDSGNSGDSPESSPEKSAPTNT"	
misc_difference		869	
		/gene="Atlg80300"	
		/note="compared to genomic sequence resulting in an amino acid sequence difference"	
		/replace="C"	
3'UTR		1903. .2151	
		/gene="Atlg80300"	
misc_difference		2134	
		/gene="Atlg80300"	
		/note="compared to genomic sequence"	
		/replace="C"	
BASE COUNT		553 a 451 c 510 g 637 t	
ORIGIN			
Query Match 21.4%; Score 350.8; DB 8; Length 2151;			







Db 1697 AACCATTTGGGGAAATGAGGGGGCTCTAAATCCAGCAGTTCATCATCTTACATTCGCG 1756  
QY 1440 AGATTTGAGGAGTATGACCCCTTATCTTCAGTATCTTTTCATCATCTTCATTCG 1499  
Db 1757 TCACTCGGCAATTCACACACTTACCTTGAGTATCTCTGCTGATATAGTACCTCATG 1816  
QY 1500 TTGGTTTCTGCAA 1512  
Db 1817 TTAGCAGCAGCTA 1829

RESULT 16  
AY084374 2146 bp mRNA linear PLN 21-JUN-2002  
DEFINITION Arabidopsis thaliana clone 105967 mRNA, complete sequence.  
ACCESSION AY084374  
VERSION AY084374.1 GI:21403084  
KEYWORDS FLI\_CDNA.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis.  
1 (bases 1 to 2146)  
Haas, B.J., Voliovsky, N., Town, C.D., Troupkhan, M., Alexandrov, N., Feldmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L.  
Full-length messenger RNA sequences greatly improve genome annotation  
Genome Biol. (2002) In press  
2 (bases 1 to 2146)  
Brover, V., Troupkhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.  
Full-length cDNA from Arabidopsis thaliana unpublished  
3 (bases 1 to 2146)  
Brover, V., Troupkhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.  
Direct Submission  
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA  
This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Ler ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. GenSet carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.  
Location/Qualifiers  
1. 2146  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/clone="105967"  
113. 1969  
/codon\_start=1  
/product="adenine nucleotide translocase, putative"  
/protein\_id="AAM60955.1"  
/db\_xref="GI:21536623"  
/translation="MEGLIOTRGILSLPAKPIVVRRLQPSHELKORLPTTNLPAISLSSGHRKFOAFQGIPLGISVSHERSGFLCKAEAAAGGANEDEGDAAAMAVSKIFGEVETLKKIKPLQVGLMEFCILFNNTILRPTKDLVLTAAAGSSAIEIPFKTWNLPMALGFMPLYTKLSLVLSKALFYIVPEIYFPGAFVYMLNLHDEKAFPLDLATLGRPMPLALIMRISRCFLPYVMALMGSVVSVLPMGCFPNOLITTDENKAFPLDGLGANVALIFSGRIVKYFSNMKNLNGPGVDGNAVSLKAMSLIVMGALICHLVWVNRV

VLEPFRSKRRKRVQWGMETMESLEFLVSSPYIRDLATLVAVGISINLEVTWMSKLA  
QFSPNENYSAEMGFSTCTGIAFTYMLLSQVYFKYKGMVAAKITPTVLLTGVAFF  
SLIFGFPAPALVQKLMPTPLAAVYVLAQVYKSKAKSLDFPCKEAVIPLDEDT  
KVKGKALIDVANCNLEKSGGALLDQOEHLILFEGSLANSTPLGIILGIVAMIAAAS  
LEOFNLMBEELEREMERASSKIPVISOEDAPSETTSOLSEKSTPGIT"  
BASE COUNT 533 a 464 c 499 g 650 t  
ORIGIN  
Query Match 20.6%; Score 336.6; DB 8; Length 2146;  
Best Local Similarity 54.0%; Pred. No. 8e-70;  
Matches 752; Conservative 0; Mismatches 614; Indels 27; Gaps 2;  
QY 120 CTAAAGAAATGTCGCCAATGTCCTAATGTCCTGTATTTAATTAAGTACGCTG 179  
Db 440 CTGAAGAAGATGTCCTTATGAGCTAATGTCCTTTCATCTTCATTAATTAACAAATC 499  
QY 180 TTACGCGATACAAAAGACACTTATGTGGAGACTCTCGTGTCTGTCGACAGGCATA 239  
Db 500 CTAGGGACACGAAGAGATGTTGTGTGAGCGCTAAAGGAAGTCTGCTGAATTTATA 559  
QY 240 CTTTCATCAAGTTTGTGCTTGTGTCCTGTCCTATATATCTTATGCTTATATGCA 299  
Db 560 CCGTTTGAAGACATGGGTGAATCTTCCGATGGCTATGGGTTATGTTCTATACACC 619  
QY 300 AAGCTAATAATTTTATAGTAAAGAGCCCTTATTTATGACAGTGGAGACCCCTTTTAA 359  
Db 620 AAACCTTCCAAATGTCCTCTCCAAAAGCCTCTTTTACATCTTATATGTTCTTTCAT 679  
QY 360 ATTTCTTGTGCGCTGTCCTCCGACGTATATTCGCGACGATGTTTATCATCTGACA 419  
Db 680 GTCTACTTTGAGAGCTTTGTTGTTGCTGATGACCTCTCAGCAATTTGATTCATCTGAA 739  
QY 420 GAATTTGCTGACGCTTACAGAGCCATCTTACAGAGTTGTAGAGCTTGTGCCATC 479  
Db 740 GCTTGTGCTGAATGATCTTCTGCAACACTCGGCCAGATTCATGAGGCTCTGCAATC 799  
QY 480 TTAAAGAACTGCAATTTGCTGCAATTTATGTAATGCTGTAACATATGGGAAGCTATG 539  
Db 800 ATGAGGATTTGAGATTTCTGTTTCTATGTCATGCTGAGCTTGTGGGATGAGTGTGC 859  
QY 540 CTATCTCTAATGTCCTGGAGATTTGCTAATGAATTTACAAATATCCAGGAACACGCT 599  
Db 860 GTTTCAGTCTCTCTCTGCGGATTTGCAACACAGATTTACACTGTGACGAAGCCAAAAG 919  
QY 600 TTCTACGCTTTTTCGGATGACGAGCTAATATTTCTTACTAGCTTGTGCTGCAAT 659  
Db 920 TTTATCTCTGTTTGGACCTTGGGCAATGTGCACATCTCTCGAGGAAGACGTG 979  
QY 660 GTTGGGCTTAAAGTGAAGCTTCCGTTTCTGAAGGTAGATCTTGGGGAATTTCT 719  
Db 980 AATATTTCTCTAATATGAGAAAGAAATCTTGCTCGAGATTTGAGCTGGGCTGTTTCA 1039  
QY 720 TTACGCTCTTGTGAGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 779  
Db 1040 TTTAAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1099  
QY 780 TGATCATATAAGACGATTTGACCGATCTCGCTTATATATTCAGAAAGATGCAAAAG 839  
Db 1100 TGGGTGATA-----GATATGAGCCCTCCCAACCCGTAGC 1135  
QY 840 GGGAAAAAGGTGCTAAACCTAAATGAATATGAACATAGCTTCTATCTTGATAGA 899  
Db 1136 AAGAAAGAAAGGTGAACACACAGATGGAACATGAGAGCTGAAGTTCTTGCTGCTCA 1195  
QY 900 TCTCTATATATTTCTTATTAATGCTCTGTGATGATGATGATGATGATGATGATGAT 959  
Db 1196 TCACCATATCATTTAGGATCTTGTACTTTGTGTGCTTCATATGATTAATCAACCTT 1255  
QY 960 ATCGAATGATCTGGAAGATGACCTGAACCTGCAATATGATATATGATGATGATAGT 1019  
Db 1256 GTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1315  
QY 1020 GAGTTATGAGGAACTTCTCTCTGAGCTGCGGTAGTATCCGACTTATCATGCTATTT 1079

Db 1316 GCATTATAGGGGAGCTTCACACCTGCACAGGATTTGGCAAC--ATTCACATGATGCTT 1372  
 Oy 1080 GTTGGTGAAGCATGATTCGTAATTTGGATGGTTAACTGAGGCCCTAGTCCTCTGTC 1139  
 Db 1373 CTAAACCAATAGTGTATTAAGATATGTTGGGAGTAGCTGCAAAAGATCCACCAACC 1432  
 Oy 1140 ATGGTCTCTCAAGAGATATCTTTCTCGCTCTTGTATCTTTAGAAACCAAGCTTCT 1199  
 Db 1433 GTTCTGATTTAGCCGGGTGCTTCTCTCTGATACCTGTTGGTGGCCCATTTCCGA 1492  
 Oy 1200 GGGCTGCTGCTATGTTGGTACAACTCCTCATGCTAGCTGTGTTGTCGAGCTATA 1259  
 Db 1493 CCATTGGTTGCCAGCTGATGACACCGTACTGCGAGCAGTGTAGCTTGTGGCCCTC 1552  
 Oy 1260 CAGATATTTCTTCCAAATCCCAAAATAGCGTCTCTTTGACTCACTAAGAAATGGCC 1319  
 Db 1553 CAGAAATATCTTACAGAGAGTGCACAAATACAGCTTGTGATCTTGCAGAAAGATGGCT 1612  
 Oy 1320 TATATCCCTCTTGACCAAGAGCAAAAGTCAAAAGTAAAGGCTGATTTGATGTAGTGGC 1379  
 Db 1613 TATATCCCTCTTGATGAGAGACCAAGGTTAAAGCAAGCTGCATTTGATGTGGTCTGC 1672  
 Oy 1380 GCCCGCTTCGAAATTCAGAGAGAGCTTTATCAACAAGTTCCTGCTTATCTGTGA 1439  
 Db 1673 AACCATTTGGGGAATTCAGAGCGGTGCTTATCCAGCACTGATCATCTTCAATTTGGC 1732  
 Oy 1440 AGTATGAGCTATGACCCCTTATCTTGATGATCTTCTTTTATATCTTCTATTTGG 1499  
 Db 1733 TCACTCGGCAATTCACACACTTACCTTGGAGTATCTGCTGGATAGTACTGATG 1792  
 Oy 1500 TTGTTTCTGCAA 1512  
 Db 1793 TTAGCAGCAGCTA 1805

RESULT 17  
 AX098422 1770 bp mRNA linear PAT 02-APR-2001  
 LOCUS AX098422  
 DEFINITION Sequence 1 from Patent WO0120009.  
 ACCESSION AX098422  
 VERSION AX098422.1 GI:13537714  
 KEYWORDS

## SOURCE

ORGANISM Arabidopsis thaliana  
 thale cress.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 1770)  
 REFINED: A., Geisenberger, P. L., Neuhaus, H. E., Graeve-Kampfenkel, K.,  
 Moehlmann, T. and Tjaden, J.  
 TITLE Plants having altered amino acid contents and method for the  
 production thereof  
 JOURNAL Patent: WO 0120009-A 1 22-MAR-2001;  
 BASF AKTIENGESELLSCHAFT (DE)  
 FEATURES  
 SOURCE location/Qualifiers  
 1..1770  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"

BASE COUNT 441 a 383 c 431 g 515 t  
 ORIGIN

Query Match 20.2%; Score 330; DB 6; Length 1770;  
 Best Local Similarity 54.5%; Pred. No. 3,1e-68;  
 Matches 769; Conservative 0; Mismatches 610; Indels 33; Gaps 4;

Oy 121 TAAAGAAAGTCTGCCAATGTTCTTAATCTTCTGATTAATTAATTAAGGCTGT 180  
 Db 335 TGAAGAAAGTATCCCTTATAGATGATGTCTTTTGTATCTTCTTCAATTCACATATC 394  
 Oy 181 TACGCGATACAAAGACACTTATTTGAGAGCTCTGGTTCTGTGAGAGGCAATAC 240  
 Db 395 TGAGGATATCAAGAGATGCTGTGTGTGACGCGCAAGAGATTTCTGCTGAGATTAATAC 454

Oy 241 CTTTCATCAAGTTTGGCTTGTGTCCTGTCGCTATTAATCAATTAATTAATTAAGCA 300  
 Db 455 CTTTCTTGAAGCTTGGGGAATCTTCTCTATGAGCCATGGTTATAGCTCTCTACACTA 514  
 Oy 301 AGCTAAGTAATATTTTAAAGTAAGCAGGCTTATTTTATGACATGGGAAGCCCTTTTAA 360  
 Db 515 AACCTCCAAATGCTCTCCCAAGAAAGGCTGTTTATACAGCTTATTTGCTCTTACCA 574  
 Oy 361 TTTTCTTGGCCCTGTTCCGAGCTAATTTATCCCTACGAGATTTTACATCTACAG 420  
 Db 575 TCTACTTTGGGGCTTTGGTTTGTCAATGATCCCTCAGCAACTATATTCACCCGGA 634  
 Oy 421 AATTGTCGACGCTTTACAGGCACTCACTACCTCCAGGATTTGTAGAGCTGTTGCATCT 480  
 Db 635 CTCTGCGAGATTAAGCTCTTACACCCCTGCGCCCAAGATTTATGAGGCTCTATTTGAAT 694  
 Oy 481 TAAGAACTGACATTTGCTGCAATTTATATGATCTGCTGAACTATGAGGGAAGCTCATGC 540  
 Db 695 TCGGATTTGGAGTTTCTGTTTATTTATGATAGGCTGAGCTTTGGGTAGTGTGTGG 754  
 Oy 541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAATTCACAGCAAGCAAGCTT 600  
 Db 755 TCTCAGTCTCTTCTGGGCTTGTCTTAATCAAGATCAAACTGTGATGAAGCCAGAAAT 814  
 Oy 601 TCTAGGCTCTTGGGATCGGATATTTCTTTACTACTCTGCTGCTGCAATG 660  
 Db 815 TCTATCTTGTGGGATGAGCAATGTGCACTGATTTTCTGAGAAACCGCTGA 874  
 Oy 661 TTTGGGCTTCAAAAGTTGAGAGCTTCCGTTTCAAGGATGATGATCTTGGGGAATTTCT 720  
 Db 875 AATACTTCTCAACTTGAAGAAAGATCTTGTGCTGGAATGA-----CGGCAAGTTTCT 929  
 Oy 721 TACGCTTTTGAATGCGTATGACTATTTATGACTGTGTTCTTATGGCACTTACTGT 780  
 Db 930 TGAAGCCATGATGAGCACTTGTGGGAAATGGAGCTCG--CATTTGCTCTCTATTTGT 987  
 Oy 781 GATCAATTAAGAGATTTGAGCAGATCCCGCTTCTATATCCAGAAAGATGCAAAAG 840  
 Db 988 GGGTGAATAGA-----TATGTTCTCTTCCACCCCGTAGCA 1024  
 Oy 841 GGAAAAAGGCTTAACCTTAATAATGAATATGAAGATAGCTTCTATCTTGTATGAT 900  
 Db 1025 AGAACAGAGAGAGAAACCGAAGATGAGAGCATGGAAGCTTGAAGTTCTGTGATCAT 1084  
 Oy 901 CTCCTTAATTTCTTTTATTAATCTCTTGTGTTATTTGCTATGATTTGCATTAACCTTA 960  
 Db 1085 CACCATATCTTAGATATCTTCTACTTAAAGTGTGAGCATGCGTATTAATCAATCTTG 1144  
 Oy 961 TCGAAGTGAATGGAAGATCAAGTCAAGTCAATATCCATATGATGATGATGATG 1020  
 Db 1145 TCGAAGTCAATGAGAAATCAAGCTTAAGCTCACTGCTCCGCGGGAATGAGTCAAG 1204  
 Oy 1021 AGTTATGAGGAACTTCTCTCTGAGAGCGGTATGATCCCTACTTAATCAATGATTTG 1080  
 Db 1205 CATTTATGAGAGCATTTCTTAACCTGACGGGTGTGCAACATTCACATGATGCT---TTC 1261  
 Oy 1081 TTGTTGATACGTCATTCGTAAATTTGATGATTAAGTGAAGCCCTAGTCACTCTGTCA 1140  
 Db 1262 TCAGCAATACGATTAATCAATAGTATGTTGGGAGATGAGCTCAAAAGATCACCCCAAG 1321  
 Oy 1144 TGGTCTCTTACAGGATATGCTTCTGCTGCTCTGTATATCTTTAGAAACCAAGCTTCTG 1200  
 Db 1322 TTCTGCTATTTAGCTGTGTGGCTTCTCTCTCTATATATTTGTTGGCCCATTTGCGAC 1381  
 Oy 1201 GCGTGGTGCATATGTTGCTGCAACTCTCTATCTAGCTGTGTTGGAGGATATAC 1260  
 Db 1382 CACTTGTGTCACAGCTGTGATATGACACCGCTACTGCTGATGTATGGTGGCCCTTC 1441  
 Oy 1261 AGAATATTTCTTGAATTCACAAATAGCGCTCTTTGACTCAACTAAGAAATGGCT 1320  
 Db 1442 AGAATATCTTACAGCAAGATGCGCAAGTACAGCTGTGTCACCCCTTGAAGAAATGGCT 1501



QY	1321	ATATCCCTCTTGACACAGACGCAAAAAGTCAAAGTAAGCGCTGCTATGTAGTGGCCG	1380
Db	1502	ATATCCCATTTGGATGAGGACACCAAGGTTAAAGGCAAAACGTCGATTCAGCTGCTGCA	1561
QY	1381	CCGCGCTGGGAAATACAGAGAGCGTTTAAATCCAAACAGCTTGGCTGCTATTCGGGAA	1440
Db	1562	ACCCATTAGGAAATACAGAGAGCGTTTAAATCCAAACAGCTTGGCTGCTATTCGGTGGAT	1621
QY	1441	GTAATGGAGCTATGACCCCTTATCTTGACAGTAATCTCTTTTCATCATTCGCTATTTGGT	1500
Db	1622	CACATGACGAATTCACACCGCGTATCTAGGAATGATCTGTTGATTAATTCACCTGGCTGGT	1681
QY	1501	TGCTTTCTGCAACTAAGTTAAACAAACTATTC	1532
Db	1682	TAGCTGACAGTAAAGCTGCTGGAGGACAGTTC	1713
RESULT 18			
LOCUS	AX100499	1823 bp	DNA linear PAT 10-APR-2001
DEFINITION	Sequence 1 from Patent WO0121803.		
ACCESSION	AX100499		
VERSION	AX100499.1	GI:13619512	
KEYWORDS			
SOURCE	Chlamydia pneumoniae.		
ORGANISM	Chlamydia pneumoniae		
REFERENCE	1 (bases 1 to 1823)		
AUTHORS	Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.		
TITLE	I Chlamydia /I antigens and corresponding dna fragments and uses thereof		
JOURNAL	Patent: WO 0121803-A 1 29-MAR-2001;		
FEATURES	Aventis Pasteur Limited (CA)		
source	location/Qualifiers		
	1..1823		
CDS	/organism="Chlamydia pneumoniae"		
	/db_xref="taxon:83538"		
	101..1723		
	/note="unnamed protein product"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAC36447.1"		
	/db_xref="GI:13619512"		
	/translation="MQSSEVRFPSRLRAYLCPIYKSEKFPVLLFAFYGVNYCLL		
	KNMEDTIVVSDAGAEVITPELKWGIVPGAVITVYVWGLSGRPYEDTVEYCEMAAF		
	LGPEFLPAVITVYVSDAGAEVITPELKWGIVPGAVITVYVWGLSGRPYEDTVEYCEMAAF		
	SVVSLMFLMGALNITTTTEAGREYALINLSNLSICAGISVWVKQFVAFSPAC		
	DSWWSLMLTGLTITTCGSLIMIMVXPRHITHTDTPSPSRVLADEGATANIKERKK		
	KPKRKARNRPHLIHOSRLRLGLATITVSYNIVHLEFVWMDVQSYTSSHVENGTW		
	SRITTLIGVSVLAVALTLTGOCIRKRWGIVPGAVITVYVWGLSGRPYEDTVEYCEMAAF		
	IFPGVGLMTPLAALMTGGMONVLRSGTKFTFEPQTEMAFIPLSPDKNHGKAIDGDIS		
	IVSVISKSGSLIQGLLIVPSSVAASLNTVAIVLIIIMVIAVIVYIGKEYYSRAAA		
	DAVATLKOPKPSISIVREAGSEVSEOMAVL"		
BASE COUNT	450 a 359 c 396 g 618 t		
ORIGIN			
Query Match	18.4%;	Score 301.4;	DB 6;
Best Local Similarity	50.3%;	Pred. No. 2.1e-61;	
Matches 814;	Conservative 0;	Mismatches 776;	Indels 27;
		Gaps 2;	
QY	2	AAATTAATACTATCAGATATGAAAATTAAGATATTCAGAGGCTAATATGACAAAAAC	61
Db	52	AACTTAATGTCTATTAAGCTTTATACGCCATTAAGATATTCAGAGGCTAATATGACAAAAAC	111
QY	62	CGAAGAAAAACCTTTTGGAAATTCGCGCTTTCTTGTTGGCCGATACATACTACGAGCT	121
Db	112	AGAAGTGAACCCCTTTTAAAGGCTCGCGGCAATCTTTGTCTCTATTTTAATGAGAAAT	171
QY	122	AAAGAAAGTCTGCGCAATGTTCTCTATCTTCTGATTAATCAATTAATACAGCGTGT	181
Db	172	TTCTAAGTTTGTTCACATATTTCTACTAGCGTTTTTTCGTTGACCTTAATCTAGCGCTGT	231
QY	182	ACGCGATCAAAAGACACTCTTATTTGTTGGAGACCTCGTGTCTGGTGCAGAGGCAATAC	241

Db	232	GAAGAAACATGAAGAGTACTGCTGATCTGTCATGTCGGTTGCAGATGCTGGCGCAACAACTGATTC	291
Qy	242	TTTTCAATCAAGTTTTGGCTGTTGTGTCCTCGTCGATATATCTTAATGCTTATTTATATGAA	301
Db	292	CTTCTCTTAAGTTTTGGGAATTTGTCGGGGAGCTGTATATGTATACATATGGCTATGGTG	351
Qy	302	GCTAAGTAAATATTTTAAGTAAGCAGGCCCTATTTTATGACAGTGGGAAGCGCCCTTTTAAAT	361
Db	352	GTTAAGCAGATCGATTCCTCGGGATACCGTTTTTATTTGCTCATGGCCCATTTCTTG	411
Qy	362	TTTTCTTGGCCCTGTTCCGCACTGTAATTTATCCGTAACGGAGTGTTTACATCTTACAGA	421
Db	412	TTTTTCTTCCCTGTTGGCTGTGATCATTTTATCCGTAGGGGAATAGCCTGCATCTCAACTC	471
Qy	422	ATTCTGACCGGTTTACAGGGCATCTACCCGAGATTCGCTAGAGACTGCTGGCATCTT	481
Db	472	TTCTCCCTGATAAATTACAGAGCTCTCTTCCAGAGCACTTCGTGGTTTTATTTGTATGGT	531
Qy	482	AAGAAACTGAGCATTTGCTGCATTTTATGTACTGTCTGAACTATAGGGAGCCTCACTGCT	541
Db	532	CCGTTACTGAGGATTACAGATATTTATTCAGTAAATGTCAGACCTGAGAGTTGCGTTTCT	591
Qy	542	ATCTCTAATGTTCTGEGGATTTGCTAAATGAATTACAAAAATCCAGAACCAAGCGTTT	601
Db	592	TTTCGATGTTGTTCGGGAGCAGCAATCAGATTACATCAATTAAGAACCGGCGCGTTT	651
Qy	602	CTACGCTCTTTTTCGGTATAGAGGCTAATTTTCTTTACTAGTCTTGCTGCTGCATTTGT	661
Db	652	TTAGCCTCTTTCAATACAGGATTAATCTCTCTCAATATGGCAGGGAAATCTCCGA	711
Qy	662	TTGGGCTTCAAGATTTGAGAGCTTCGGTTTCTGAAAGGTATGATCTTGGGGAAATTTCTT	721
Db	712	TTTGGATGGGGAACAACATTTGTTGGCTACTCCTTTGCATGTGATTTCCGTGGACTCTGT	771
Qy	722	ACGTTCTTTGATGGCTATGACTAATTTATCTGCACTTGTTCTTAATGGCAGTTACTGGTG	781
Db	772	AATGCTCAACTTGACATAGCTGATCACTTGTCTGGTTTATTAATGATCTGGCTATATAG	831
Qy	782	GATCAATAGAAGCATTTGACCCGATTCCTCGCTTC-----TA	817
Db	832	GCGGATTAATCATTTTGACTATGTGATTTGATTCCTCCATCTAGACGTGTCTTGGCAGA	891
Qy	818	TATTCAGAGAAATGCAAAAGGGGAAAAAGGTGCTAAACCTAATAATGAATATGACAGA	877
Db	892	AGAGGGACACTCTGCTATATCTAAAGGAAAAAGAAAAACCTAAGGCCAAGACTAGAA	951
Qy	878	TAGCTTCTCTATCTGTGATGATCTCCTTAATATCTTTATTAACGTCCTTGCTATTTGC	937
Db	952	CTTTTCTTACACCTCATTCAGTCTGCTTATTTATTAAGGGCTGCTATTATTTGCTATC	1011
Qy	938	CTATGGTATTTTGCATTAATCACTTAATCGAAGAGCACTGGGAAAGTCCACTGAACTGCAATA	997
Db	1012	CTATATATTTGGTGATTCATCTATTTGCAAGTCGTTTGGAGAGATCAAGTTAAGCCAGATTTA	1077
Qy	998	TCCTAATATGAATGACATATAGTAGTTCATAGGGAACTTCTCCTTCTGCACTGGCTACT	1057
Db	1072	CAGTTCTCAGCTAGAAATTCATATGGGTATAGAGATAGATCACTAACCTCATTTGGGTGCT	1133
Qy	1058	ATCCGTAATTAATCATGCTATTTGTTGGTGTAAACGTCATTTGTAATTTTGGATGCTTAAC	1117
Db	1132	TTTCTGATTTACACCTGACTACTCCTTACCGGACAGTATATCCGTAATATGGGAGTAGCACT	1197
Qy	1118	TGAGACCTATATCACTCTGTCATAGGTCTTCCTCAACAGATATGCTTTCTTGCTCTTGT	1177
Db	1192	CGGTGCTTTTACTCTCATCTCATTTGGTAATAGTTTACAGACATGCTCTTTTGGAACTAT	1257
Qy	1178	TATCTTTTGAAGAACCACT---TCTGGGCTGGGCTGATGTTGCGTACATCTCTCAT	1234
Db	1252	TTTTTCTGCAAAAGAGACATCTCTATTTTGGGGAGTTCTTGGATATACACCTCTGGC	1311
Qy	1235	GCTAGCTGTGGTTGTCGGAGCTATACAGATATTTCTTTCGAAATCCACAAATATAGCTCT	1294



Db 1312 TCTAGCTGCTGGAGCTGGAGGATGCMAAATGCTCTATCCCGGGGAGCGAATTTACGTT 1371  
QY 1295 CTTTACACCACTAAAGAAATGGCCATATCCCTCTGTGACCAAGCAAAAGTCAANG 1354  
Db 1372 CTTTATCAAAACCAAGAAATGGCCTTATCCACTTCTCCAGAGGATAAATATCANG 1431  
QY 1355 TAAGCTCTATTGATGTAGTTGACCGCCGCTTGGAAAATCAGAGAGGATTTATCCA 1414  
Db 1432 GAACCCCGGATGTATGGTGTGTTCAAGGATAGSAAAGTCTGGAGGCTCTTTATTTA 1491  
QY 1415 ACAAGTTTGTCTGTATCTGTGGAAGTATGAGCTATGACCCCTTATCTTGACAGTGAT 1474  
Db 1492 CCAAAGGCTGCTGTATTTCTCTCTGTCACAGCAAGTTTAAACGTCATCGCCCTAGT 1551  
QY 1475 TCTCTTTTTCATCTTGCATTTGTTGGTTTCTTCCAACTAGTTAAACAACTATCTT 1534  
Db 1552 TCTCTCATATATGATGCTGTTGATGCGGTTGTGCTTATTCGTAAGAATACTTA 1611  
QY 1535 AGCGCAGTCTGCTTTAAAGAACAAAGTGCCTCAGAAAGTTCAGCTCTGCTTC 1591  
Db 1612 CTTAGAGCTGCTGATGCTGATGCAACCTTGAACACCTAAAGAACCTTCTCTTC 1668

RESULT 19  
AE001646 10163 bp DNA linear BCT 01-DEC-2000  
LOCUS Chlamydia pneumoniae section 62 of 103 of the complete genome.  
DEFINITION AE001646 AE001363  
ACCESSION AE001646.1 GI:4376910  
VERSION AE001646.1 GI:4376910  
KEYWORDS  
SOURCE  
ORGANISM Chlamydia pneumoniae CWL029.  
REFERENCE 1 (bases 1 to 10163)  
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PubMed 10192388  
REFERENCE 2 (bases 1 to 10163)  
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
FEATURES  
source location/Qualifiers  
1. .10163  
/organism="Chlamydia pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
complement(93. .1715)  
/gene="adt.2"  
/note="Cpn0614"  
complement(93. .1715)  
/gene="adt.2"  
/codon\_start=1  
/transl\_table=11  
/product="ADP/ATP translocase"  
/protein\_id="AAD18753.1"  
/db\_xref="GI:4376911"

gene  
CDS  
complement(1944. .2450)  
/gene="pgsa\_1"

CDS  
/note="Cpn0615"  
complement(1944. .2450)  
/gene="pgsa\_1"  
/codon\_start=1  
/transl\_table=11  
/product="Glycerol-3-P Phosphatidytransferase"  
/protein\_id="AAD18754.1"  
/db\_xref="GI:4376912"  
/translation="MROPCLNLSRLMLALFYOEKIHILALIVGAMLSVDYDGL  
ARRKATSLGSLIDPTDKVEFVCTIVLYMESLSIAHLFFICARDLFLIFCYL  
SLVGMKGVDYDGLFWGKIFIVVOFILLGATAGGELIPWGLVPLVGLFLEIRIM  
DYKROFLR"  
2672. .2753  
/product="tRNA-Ser"  
3104. .4453  
/gene="dnab"  
/note="Cpn0616"  
3104. .4453  
/gene="dnab"  
/codon\_start=1  
/transl\_table=11  
/product="Replicative DNA Helicase"  
/protein\_id="AAD18755.1"  
/db\_xref="GI:4376913"  
/translation="MIVGLMGTGVHVLNLANOLYEEDEFYLEHKIIFRYLDAFKO  
DKPIDVHLAGEBELRHNOITVIGGPSYLTITAEFAGTAAYLEEVVDIIRSKM  
STAKIERKALFQRKNVALEDEKONSFFKISQTSVSQYTLVADKLRLGTLTTDPKY  
LVQLOERLELTONAGDNKSEFFYGIPHFIDLDOLHGFSPSNLMLTAAPAMKTA  
LALNIAENLCFQNRLLPIGIFSLVDLVQDLIRHMCISREVSQKTSIDLSGHFORI  
VSVINEMOEHLLIDDPGLKSLVDLRARARMKESYDIQFLIDYLOLSSGSLRAT  
ESRQTEISEIRMLKTLARELINPILCLOSRSKVEDRANRPMMSDLRESGTEODS  
DLVMPLEIREVYDNDKGTAEMLIIAKNRHSISVPLVFEKELARPNSAFETIS"  
4797. .6632  
/gene="gida"  
/note="Cpn0617"  
4797. .6632  
/gene="gida"  
/codon\_start=1  
/transl\_table=11  
/product="FAD-dependent oxidoreductase"  
/protein\_id="AAD18756.1"  
/db\_xref="GI:4376914"  
/translation="MMTHPIAVDYIVVAGHAGEAAACSAKMGVSLMLNSLDTIA  
KLSCPNPAVGIGKHIVREIDALGIMAEVVDOSGIORILNORKGPAVRAPRNOVK  
OLYTHMKRLLENPGHIMATYBSLDKRGYISGYTTKGMFSGTYVLSGCTPM  
RGLIHIDGNFSGRLGDPSSQSLSEDLKRGFTSRKTGTPRELASSINFCMEB  
QPGDVGVEVHRTPEFQPLPOLSCFTITMEKKAIIISANLHRSALYGCIEVGPR  
YCPSEIDKIVKSDKERHVFLEPGLTOEIVANGISTMPFVOYDMIRSVGLEIN  
AITRPAVAIEVDYIHGNVIHPTLESKLEGLFCGQINGTGGYEAAAGLIGINA  
VNKVENRPFIPSRQESYIGVLDLITQOIDEPRYRMPYGRAEHRLIRODNACARLS  
HYGEIIGLSEERYEIVKQNDLEEVYRLOKPEYRQVGSVSLAKASISPEYSYDM  
LREAFPNDIRLCAVLNANSLMEIKYSYIDOKILITSLSKASLLLPEDLDYKQIT  
ALSLAEQEKLAFTPRPLGSASRISGASADIOYMLAKRHHH"  
6619. .7326  
/gene="ip1A\_2"  
/note="Cpn0618"  
6619. .7326  
/gene="ip1A\_2"  
/codon\_start=1  
/transl\_table=11  
/product="Lipote-Protein Ligase A"  
/protein\_id="AAD18757.1"  
/db\_xref="GI:4376915"  
/translation="MPTTNCIFLDIRGHSLIHQIOTIEALLRVANONFCIINSGAKOS  
IVLGSLRNLDNDVHISRAQADHIP1IRYSGGVVFIDSNLAWSWINNSASAQO  
ELAWTVIYSPDLNPTFSIRENDYVIGHKIGGNAOYIOHRVWHHTTFPLMDLTK  
LSYVLPFQOQPTVRNORSHHEFLTLRPWPSDDFLERIKASGSLFTWEFLONE  
LELILAOHRKATVNLN"  
complement(7323. .7757)  
/gene="ndk"  
/note="Cpn0619"  
complement(7323. .7757)  
/gene="ndk"

212 TTTCATCAAGTTTGGCTGTGTCCTGCTGCTATTATCTTATATGCTTATTTATGCAAA 301  
 1524 CTTCCTTAAGGTTTGGGAATTTGCCGGAGACTGTATTGTACTATAGCTATGGGTG 1465  
 302 GCTAGTAAATATTTAAGTAAGCAGAGCTTATTTATGCAAGTGGGAAGCCCTTTTAAT 361  
 1464 GTTAGCAGAGTGGATATCCCGGGATACCGTTTATTTATGCTTACATGGCCGATCCCTTG 1405  
 362 TTTCTTTCCCGTGTCCCGACTGTAAATTTATCCGCTACGCACTATTTTACATCCCTACAGA 421  
 1404 TTTTTCCTCTCTGTTTGGTGTGATCATTTTATCCGTAGGGATATAGCTGCATCTCAACTC 1345  
 422 ATTTCGTACCGTTTACAGGCCATCTTACCTCCAGAGATTTGCTAGAGCTGTTGCCATCTT 481  
 1344 TCTCGCTGATAAATTTACAAAGAGCTCCCTCCCAAGGACTTGCTGTATTATTTGATGGT 1285  
 482 AAGAAATCGGACATTTGCTGCATTTTATGTACTCTTGCACTATAGGGAAAGCCATCATCT 541  
 1284 CCGTACCTGGAGTTACAGATATTTATTAGCTATGTCAAGAGCTGGAGTTGGTGTCT 1225  
 542 ATCTATATGTTTTCGGGATTTGCTAATGAATTAACAAATTCACAGACAGCAAAAGCCTTT 601  
 1224 TTTCATGTGTTTCTTGGGACTAGCCAAATCAGATTACTACAAATTTACTAAGCGGCCCTTT 1165  
 602 CTACGCTCTTTTTCGGTATCGAGGCTAAATATTTCTTACTAGCTTCTGTGCTGCATTTGT 661  
 1164 TTACGCTCTTATCATATACAGATTAATCTCTCCATATAGCCAGAGAAATCTCTTA 1105  
 662 TTGGGCTTCAAAAGTTGAGAGCTCCGTTTCTGAAAGTGATGATCTTTGGGAATTTCTTT 721  
 1104 TTGGATGGGGAACAACATTTGTTGGCTATCTCCCTTGATGATGATTCGCGGACTCTGT 1045  
 722 ACGTCTTTGATGGCTATGACTATTGTATCTCGAGCTTGTCTTATGGCCAGTTACTGTG 781  
 1044 AATGCTCACTTGACACTGCACTGATCACTTGTTCGGTTAATATGATCTGGCTATATAG 985  
 782 GATCAATTAAGACGTATTTGACCGATCTCGCTTC-----TA 817  
 984 GCGGATTCATTTTGAATATGATTAATGACTGTGATCCCTGCATCTGAGAGTGTCTTGGCAGA 925  
 818 TAATTCAGACAGAATGCAAAAGGGGAAAAAGTGCTAAACCTTAATATGAATATGAAGA 877  
 924 AGAGGAGCACCTACTGCTAATCTTAAGGAAAAAGAAAAACCTTAACCCAAAGCTTAAGAA 865  
 878 TAGGCTCTCATCTGTATAGATCTGCTAATATCTTTATTAATCTCTGTGGTTATGCG 937  
 864 CCTTTTCTTACACCTCATCTACGTCTCGTTATTTATTTAGGGCTCGCTATTTATGTCTATC 805  
 938 CTATGTTATTTGCAATTAACCTTAATTCGAAGTGAATGGAAAAAGTCAGCTGAACCTGCATA 997  
 804 CTATATATTTGGTATGCATCTATTTGCAAGTGCTTTGGAAGAGATCAATTTAGCCAGATTTA 745  
 998 TCTTAATATGATATGATATAGTAGTGAAGTTCAATGGGGAACCTTCTCTTCTGCACTGGCGTAGT 1057  
 744 CAGTTCACAGCTAAGAAATTCATAGGGTATATGAGTATGATCACTACCTCATTTGGCCGTGT 685  
 1058 ATCCGTAATTTATGATGCTATTTGTTGGTGGTAACGTATCTTAATATTTGGATGGTTAAC 1117  
 684 TTTCGTATTTAGCAGCTGACTACCTTACCCGCACTGATATCCGTAAATAGGGAGATGGACTGT 625  
 1118 TGGAGCCCTATGCTCTGTCATGATGTTCTCCCAACAGATATGCTTTTCTCTGCTTGT 1177  
 624 CCGTGCTTTTACTCACTCATTTGTAATGTTAGTTTTCAGGACTGCTCTTTTTCGGAACATAT 565  
 1178 TATCTTTAGAAACCAAGCT--TCTGGGCTGGTGGCTATGTTGGGTACAACTCTCTCAT 1234  
 564 TTTTGTGCGCAAAAGACACATCTCTATTTTGGGGGAGTCTTGGGAATGACACTCTGGC 505  
 1235 GCTAGCTGTGGTGTGGAGCTATACGAATATTTCTTGAATATCCACAATATACGCTCT 1294  
 504 TCTAGCTGCGCTGAGCTGAGGAGATGCAAAATGCTCTATCCCGGGGAGCAAAATTTACGTT 445

gene  
CDS  
/protein\_id="AAF38008.1"  
/db\_xref="GI:7189060"  
/translation="MRLSLILKLHFLSRSSSSLSPHYHSCSRSMALLLCRMKADQIMEMOQICILIEGVCSSRMGKLVLSQKFEQSCDQJQHEHRILQYREQLSALEEYRRREBAKNODKICILEGNTWLNQRLAEKTLQOI RHSDIIDEKLEKLOSQVQRTSEGRCLQYEHKIKOLEQLOLVYVSOHQDAPSTIEIEDKSSAAVAELNRKSLIDLOEKQIYITVHSEAKRIEKLQREQAGQTSSEVCSIEKLETVQDIAEKKKAILADQIDVQCYCLKDIHEKQMDPNSNKLIDHLKGLLQKEPESEVDVFSKSLGS"  
1448..1954  
/gene="CP0126"  
1448..1954  
/gene="CP0126"  
/note="similar to GP:999741; identified by sequence similarity; putative"  
/codon\_start=1  
/product="Holliday junction resolvase"  
/protein\_id="AAF38009.1"  
/db\_xref="GI:7189061"  
/translation="MSELIGVDPGTIVAGYAIIAIVQRYQLRPYSQAIRLSSDMPULPMRYKTLFEQLSGVLDTPQANVALEQDFVKNQDSQTMKLAMARGIIVLLAAQGDILLIEYAVANVAKAVGKGHASKRQGVQVMSKILNPEVLHPSMEDIAQAFALACHTHVAARSLQGVH"  
1956..2579  
/gene="CP0127"  
1956..2579  
/gene="CP0127"  
/note="similar to PID:1183841 SP:Q51425; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="Holliday junction DNA helicase RuvA"  
/protein\_id="AAF38010.1"  
/db\_xref="GI:7189062"  
/translation="MYDRIKGTITFYVHNGAIVICQGIYAIATERPAICPAICLAIHQVDELVAITHIPEETELHLKGFHSRREBEFRLLIFSGIGKRLAIIALNALPRLKVLQVVRSEDIRLASVSGIGKTKTAKLWELKQKLDLLPLDSRVTQTPHTSSCLEEGIQALAAAGYKRIAAERMIAEIKDLPESSLDILCPALKKNSGVNKD"  
2651..3085  
/gene="CP0128"  
2651..3085  
/gene="CP0128"  
/note="similar to SP:P24233 GB:X57555 PID:416172 GB:U00966 PID:1788866; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="nucleoside diphosphate kinase"  
/protein\_id="AAF38011.1"  
/db\_xref="GI:7189063"  
/translation="MEQRLSIIRKPDVSKAHIGETLSIFEQSGRIAMKMHLSQTEAEGFEYFHRERPEQELVDPMVSGPVVVLVEGANVANSRNEMLCATNPDAAGCTIRRAKFGESIVNVAHGSDLENAVAIEAIVFFSKIEVYVNSKPLV"  
3087..3789  
/gene="CP0129"  
3087..3789  
/gene="CP0129"  
/note="similar to GB:U14003 SP:P32099 GB:X03046 PID:432654 PID:504496; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="lipote-protein ligase-related protein"  
/protein\_id="AAF38012.1"  
/db\_xref="GI:7189064"  
/translation="MPTNCTFPIDLRGSHLHQIIEPALLRVANONPCINSAGKQSVLVSGRINQDVHISRAQDHPITIRYSSGGVTFDVSNTLNAISWLNSSASAKQSPQELLAWTYGISPLPNTFSIRENDVLGKHKIGGNAQYIQHRVNHHTTFLWMDIDLQKLSLYLPIQOQGYRNQSRSHHEFLTLTRPMFPSSRDFLERIKASGLSLFTWEEFLDNELEELIAOPHRKATVVLN"  
3791..5611  
/gene="CP0130"  
3791..5611  
/gene="CP0130"  
/note="similar to GB:LI0328 SP:PI7112 GB:K00826 GB:X01631

PID:290590: identified by sequence similarity; putative"

```

/codon_start=1
/transl_table=1
/product="gidA protein"
/protein_id="AAF38013.1"
/db_xref="GI:7189065"
/translation="MHTPIAYDIYVAGACGEAAVCSAKGVSVLMTSNDITLA
KLSNPVAGGIGKGIHVEIDALGIMAEVTDQSGIQFRLNQKGAIVAPRAQVVK
QLYHMKRLIENTPGLHMQAVESLDEGVISGVTTKGMMSEKTVVLSQGTG
RGLIHIDRNESGRLGDPSSQSLSEDLKRGFPISRLKGTGPPRLASSINFGMEE
OPGDLGVGFHRTPEPQPLPOLSCFITHTEKTAIIISALHRSALYGGCIEVGR
YCPGIEEDKIVKESKERNHVEPEGLHTEIYANGSTSPREPQVDMRSVGLN
AIIIRPAIAYEDYHGNVHPTLESKLISGLFCGQINGTGTIEEAAGQSLIGINA
VNVKRNRPETIPSRQESYIGVMDLDTQLIDEPYRMFTGAERHLLRDNDACARLS
HYGEILGLSEERYELVKQNLLEBKRVLOKTFRQYQGVSVSLAKALSPEVSYDM
LREAFPNIDIDLGAVLNASLEMEIKYSYIDROKILIOSLEAESLLIPEDYKQIT
LSLEAOEKLAKEPRTIGSASRISGIASADIQVLMALKKHAHH"
complement(5955..7361)
/gene="CP0131"
/complement(5955..7361)
/gene="CP0131"
/annotation="similar to SP:P03005 GB:K01174 GB:I02312 PID:145763
PID:145765: identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="replicative DNA helicase"
/protein_id="AAF38014.1"
/db_xref="GI:7189066"
/translation="MDKSTGVLPSPHSESEMIYVLCMTGVHYVNLAAANLYEED
FYLEHKLIFRVLQDAFKODKPIDVHLAGELIKHKNQITVYGGSYLITLAEFGTAA
YLEBYDILRSKLSLRKMISTAKTEKRALEOKNVBALEAENSFFKISOSTVSQ
YTLVADLRGLTITTDKPYVLODERQLELQNAQGNKSEFTGIPHFIDLDLQING
FSPNLMILAAPAMGKTALANLAEMLCFQNRLPDIFSLMTVDQLIHMKISRE
VDSKKSISIDGLSGHFORIVSINEMQETLLIDDPGLSVDIRAARAKRESYDIO
FLIDYQLQLSSGCTLRATESRQETISEIRMLTLARELINPILICSOISRKVEDA
NHRPMMSDLRESGSIEDSDLVMLLRREYVDPDKGTAEILLIAKKRHSIGSVPLV
FEKLAERFNYSAECIS"
7654..7741
/gene="trna-Ser-1"
7654..7741
/gene="trna-Ser-1"
/product="trna-Ser"
7958..8464
/gene="CP0132"
7958..8464
/gene="CP0132"
/annotation="similar to GB:M12299 SP:P06978 PID:473749 GB:U00096
PID:1736571: identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="CDP-diacylglycerol--glycerol-3-phosphate
3-phosphatidylyltransferase protein, putative"
/protein_id="AAF38015.1"
/db_xref="GI:7189067"
/translation="MRQFCNLSLRMLALYCOEKLHRLAIYCAMISDVLDGL
ARRKATSRIGSLIDPTDKVAVFVCTIVYMBSSTIAHFFICARDLILFVCL
SLVGMKGVDYGSFLFMGKIPTVVOFILLIGVTAGGEIPMTGLPVALGLPYLERIM
DYKKOFLR"
8693..10315
/gene="CP0133"
8693..10315
/gene="CP0133"
/annotation="similar to GB:M28816 SP:P19568 PID:152470
GB:AJ235269: identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ADP, ATP carrier protein"
/protein_id="AAF38016.1"
/db_xref="GI:7189068"
/translation="MOSSEVKKPFSRLRAVLCPIYKSESKFVPELLAFVGNVCL
KNKMDLIVIGSDAGAEVLEPLKVGIVGAVIVTVYGMKSHYPRDIVECYEMAF
LGPEPLFAVIVPGDSLHNLADKIOELLPGQIRGFIWVRVWSYISIVYVSEMS
SVLSMLFWGLANDITTTTAGRFYALINTGLNLSSICAGEISITWKKOTFVAVSPAC

```

Query Match 18.4%: Score 301.4; DB 1; Length 10407;  
 Best Local Similarity 50.3%; Pred. No. 1.8e-61;  
 Matches 814; Conservative 0; Mismatches 776; Indels 27; Gaps 2;

```

2 AATTAATAAACTATACAGATGAAATAAAGTATTTCAGAGGGTAATATGACAAAAC 61
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8644 AACTTAATGCTATATAGCTTTATACGCCATTAAGTTGGAGAGGATTATGCACTATC 8703
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
62 CGAAGAAAACCTTTTGGAAAATTGCGCTCTTCTTGGCGGATACATCTACAGAGCT 121
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8704 AGAAGTAAACCCCTTTTGAAGCGCGGAGCATCTTGTCTCTATTTAATGAGAAATT 8763
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122 AAGAAAGTTCTGCCAATGTTCTCTATATGTTCTGCTATTAACATTTAATAGGTGT 181
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8764 TTCTAAGTTTCTCCACTATTTCTACAGCGTCTTCTGCGCTTAACTAGCTGCTGCT 8823
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 ACGGATCAAAAGACACTCTTATGTGGAGCTCGTCTGTCGAGAGCAATACG 241
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8824 GAAAACATGAAGATACCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8883
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 TTTCATCAAGTTTGGCTTTGTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8884 CTTCCTTAAGTTTGGGAAATTTGCCGGAGCGTGTATTTCTACTATGCTATGCTG 8943
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 GCTAAGTATATTTTAAGTAGCAGAGCTTTATTTATGCAAGTGGGAAAGCCCTTTTAT 361
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8944 GTTAGGCGAGTGGATGATCTCGGGAATACGTTTATTTATGCTTATGCGGATCTTGG 9003
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 TTCTTTGCGCTGTTCCCGCATGTAATTTATCCCTACGCGATGTTTACATCTACAGA 421
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9004 TTTTTCCTCTGTTTGGTGTGATCATTTATTCGTGATGAGGATAGCCGCTTCACATC 9063
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 ATTTGCTGACCGTTTACAGGCCATCTACATCCAGAGATGCTAGAGCTTGCCATCTT 481
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9064 TCTCGCTGATTAATTAATACAGAGCTCTCTCAAGAGACTGCTGTTTATTTAGATGGT 9123
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
482 AAGAACTGACATTTGCTGATTTTATGTAATCTGTAACATATAGGGAAGCCATGCT 541
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9124 CCGTACTGAGATACATATTTATTTACATATGCAAGCTGAGATGCTGCTGCTGCT 9183
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
542 ATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAATCCACAGCAAGACGTTT 601
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9184 TTGCATGTTGTTCTGGGAGTACCAATGACATGACATTAATGATGAGGCGCGTT 9243
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
602 CTACGCTCTTTGCGATCGAGCTAATATTTCTTACTAGCTTCTGCTGCAATGCT 661
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9244 TTACGCTCTTATCAATACAGATTAATATCTCTCATATGCGCAGAGAAATCTCCTA 9303
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
662 TTGGGCTTCAAGATTTGAGAGCTTCCGTTTTCGAAGGTATATCTTGGGAATTTCTT 721
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9304 TTGATGAGGGAACAAACATTTGTTGCTCTACTCTTTCGATGATATCTTCCGCACTG 9363
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
722 ACGTCTTTGATGAGCTATGATATTTGATGATGATGATGATGATGATGATGATGATGAT 781
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9364 AATGCTCACTTGAACATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 9423
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
782 GATCAATAGAACGTATTTGACGATCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 817
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9424 GCGATTCATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9483
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
818 TAAATCCAGAGAAATGCAAAAGGGGAAAAAGGTGCTAAACCTTAATGATATGAGAGA 877
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9484 AGAGGAGGACACTGCTGATATCTAATGAGAAAGAAAAAGAAAGCAAGCTTACAAA 9543
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
878 TAGTCTCTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 937
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9544 CTTTTCCTACACCTCATTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9603
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

OY	938	CTATGCGTATTGGACATACTTAAGTTAAAGAGTGCCGAAGAAGTCAGCGTAAACATGCACA	997
Dd	9604	CTATATATTGGTGATGCCATCTCATTTCCGAAGTCGTITGGGAAGATCAAGTTAGCCAGATT	9663
OY	998	TCTCAATATGAATGACTATTAAGTAGATTGATGGGCAACCTTCCTCTTGAGTGGGCTAGT	1057
Dd	9664	CAGTTCTCAGCAGATTCATCGGTATATATGATAGATTCACACTACCCTCATTTGGGCTGCT	9723
OY	1058	ATCCGACTATTCATGCTAATTTGGTGGTAGACGTCAATCGTAATTTGGATGGTTAAC	1117
Dd	9724	TTCTGTATTACAGCTGCTACTCTTACCAGGACAGTGTATCCGTAATATGGGATGACGTG	9783
OY	1118	TGGAGCCCTAGCACTCCTGTCATNGCTCTCCTAACAGATATCGTTTCTTCCTGCTTGT	1177
Dd	9784	CGGTGCTTTTACCTCATCTTGGTATATTTAGTTTCAGAGACTGCTCTTTTTCGGAACAT	9843
OY	1178	TATCTTTTGAAGAACCAAGCT---TCGGGCTGGTGCCTATGTTTCGGTACACCTCCGAT	1234
Dd	9844	TTTTGCTGCAAAAAGACACATCTCATATTTTGGGGATTTCTTGGAAATGACACCTCTGGC	9903
OY	1235	GCTAGCTGTGGTGTGGAGCTATACAGAAATATCTTTTGGAAATCCACAATAATACGCTCT	1294
Dd	9904	TCTAGCTGCTCGAGACTGGAGGATGCAAAATGCTCTATCCGGGGAGCAAAATTTACGTT	9963
OY	1295	CTTTGACCTACACTAAAGAAATAGGCTATATCCTCTTGACCAAGGACAAAAAGTCAAAGG	1354
Dd	9964	CTTTGATCAAAACCAAGAAATAGGCTTTATCCCATCTTCTCCAGAGGATTAATAATCATAGG	10023
OY	1355	TAAAGCTCTATATGATAGTAGTGGCCGCCGCTCGGAAATACAGAGAGGCTTTATGCCA	1414
Dd	10024	GAAGCCCGCATATGATGGTGTGTCGTTTCAAGATAGGAAAGCTTGAGAGCTCTTTAATT	10083
OY	1415	ACAAGTTTGCTCGTTATCTGTGGAAGTATTGAGACTATGACCCCTATCTTGACAGTAT	1474
Dd	10084	CCAAGGCTGCTGTTGATTTTCTCTCTGTTGCACAGATTTAAACGTCATCGCCCTAGT	10143
OY	1475	TCTTCTTTTCACTCATGCTATTTGGTGTGTTTCTGCAACTAAGTAAACAAACTATTTCTT	1534
Dd	10144	TCTTCTCATTTATTAATGATGCTTTGGATGTCGGTGTTCCTCATATATGCGGTAAAGAA	10203
OY	1535	AGCGCAGCTGCTGCTTTAAGAACCAAGATGGGCTCAACAGATTCAGTCCGCTGC	1591
Dd	10204	CTCTAGAGCTGCTCATGCTGTAGCACACCTTGAACACACCTAABAGACCTTCTCTTC	10260
RESULT 21			
AP002547/c			
LOCUS	300550 bp	DNA	linear BCT 25-MAY-2002
DEFINITION	Chlamydia pneumoniae J138 genomic DNA, complete sequence,		
ACCESSION	AP002547		
VERSION	AP002547.2		
KEYWORDS	GI:10176693		
SOURCE	Chlamydia pneumoniae J138 (strain:J138) DNA.		
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
REFERENCE	1		
AUTHORS	Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M., Takeuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H., Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A., Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T. Comparison of outer membrane protein genes omp and pmp in the whole genome sequences of Chlamydia pneumoniae isolates from Japan and the United States		
TITLE	J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)		
JOURNAL			
MEDLINE	20298986		
REFERENCE	2		
AUTHORS	Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K., Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T. Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA		
TITLE	Nucleic Acids Res. 28 (12), 2311-2314 (2000)		
JOURNAL			
MEDLINE	20330349		

REFERENCE	3 (bases 1 to 300550)
AUTHORS	Shirai, M.
TITLE	Direct Submission
JOURNAL	Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail:mshirai@pe.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227, Fax:81-836-2415)
COMMENT	On or before Sep 15, 2000 this sequence version replaced gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:6978889.
FEATURES	Location/Qualifiers
SOURCE	1. 300550
gene	/organism="Chlamydomonas reinhardtii J138"
CDS	/strain="J138"
gene	/db_xref="taxon:138677"
CDS	/note="synonym:Chlamydia pneumoniae (strain J138)"
gene	complement(21..794)
CDS	/gene="dapB"
gene	complement(21..794)
CDS	/gene="dapB"
gene	/codon_start=1
CDS	/transl_table=1
gene	/product="diaminopimelate epimerase"
CDS	/protein_id="BAA88725.1"
gene	/db_xref="GI:8978890"
CDS	/translation="MARSPSTSKYFETYSAGNRFLLGELTLPVEDVRELCQETRVDFGLKPPSCADQILITFNSDGRPTMGNGRLCAIHLASOKGSDISYSTSGLYSGYFSDRLVLDMTLADMRASVHRLSERPPLPEVVCITGVPAHVLPEISTDLSDLEPRLVHQTGSEPDGVNVEVQILGHCOLRVITYERGVEGTAAAGCTGALVALVVSNSGKKESTDIHWGGEELMVSNRNRVYIQGCVTDL"
gene	complement(763..1338)
CDS	/gene="clpP_1"
gene	complement(763..1338)
CDS	/gene="clpP_1"
gene	/codon_start=1
CDS	/transl_table=1
gene	/product="Clp protease"
CDS	/protein_id="BAA8726.1"
gene	/db_xref="GI:8978891"
CDS	/translation="MADGEVHAKRDIIEKELLEARVFESPEYKESASDAIKKWLTLERKGRPIVIVINSRPGSVANDQIKMLTSPVTVTGLAASGVLSICAPGRRTPHSRIMIHQPSITGPTTGCAQDLDIHAEILTKARIIDIVEATNOPRDIIEKALIDRMAMMTANAKPFGLDGLTFSEFND"
gene	complement(1358..2917)
CDS	/gene="glvA"
gene	complement(1358..2917)
CDS	/gene="glvA"
gene	/codon_start=1
CDS	/transl_table=1
gene	/product="serine hydroxymethyltransferase"
CDS	/protein_id="BAA98727.1"
gene	/db_xref="GI:8978892"
CDS	/translation="MLKVEKFKKFAIVEITFKVAVVSLHKLFLNASKGKQGLASTAYIALDLHLNAPSTIGERLIIDELKSKORSHLKATASENNTSLSVQAMGLDLTKGEGSPFKRRYSCEENVDAIEMECVETARLEPADACVQPSHGADANLLAWAILTHRVQGSFVSKRYGKTYVELTEETTLKAKMSCVCCGPSLNSGGLTHGNVRLNWSKTLMRCEFYDNPDECEQDAEISRLAKYPRKVLIAYSSTYRNLFAVLAQEDGSVLAWDNAHEGLTAGVAGCVFADEENRPIYADIVITTTHTKLRGPRGGLVLAAREESTLNKACPLMGSGPLPHVIAKTYALKEALSVDFPKRAHVVYNNARLARERLSHGRILTGTDNHMAHEDVGLSGISGKIADILSSVGIAVNRNSLPSDAIGKWDTSIGRTGPAITTLGNGHIDEEVADIVIKVLRNIRLSCHVEGSSKKNKGLEPAIAQEARVRNLLRPP
gene	LPYIDLEALV"
CDS	3107..3778
gene	/gene="CpJ0522"
CDS	3107..3778
gene	/gene="CpJ0522"
CDS	/codon_start=1
gene	/transl_table=1
CDS	/product="CT433 hypothetical protein"
gene	/protein_id="BAA98728.1"
CDS	/db_xref="GI:8978893"
gene	/translation="MTLVGLGNQTKARKYQAHVLPITLPEYAKSTPQNRKALQFLPQATNHLVSPSTHFLFSRMTSLSKATLKTKTYCTIGESTKRELLSFLGQGVKVAQVQ

gene  
CDS  
3843..4175  
/gene="CPj0523"  
/gene="CPj0523"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98729.1"  
/db\_xref="GI:8978894"  
/translation="MASSATPEFDCGTASLEPPATRPYMKRLAFTVIALALWIAL  
IATYTAIGLCIHPCLSFLEFLTAIPLYISRYICSHYARNVYALDNDVDHSLKIDMRS  
HSPFSDR"  
4223..5302  
/gene="CPj0524"  
4223..5302  
/gene="CPj0524"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98730.1"  
/db\_xref="GI:8978895"  
/translation="MSGPSRTSSOVSVLSVPRDKETAPKKOFTIAKISLAIASL  
ALGALVAGISLTVLGNPVFLALITLALFVSVPFLVYHOMTSKSSWMQVYLEQNEK  
PLGAMOEKNVDCYSNEMQFNNHLPKFAVIAQDASQPOPTFLNGLRYIEKQST  
GIIENPVPTNLIDNTASTLIITSTLKDKSMDTCORSGPAKGDSPSPTEVR  
VKLPHEALDQFENLNLSSAEKKSILPTFLHVCPCSELPNOOEYRQALLAENCL  
KAIESHAIVLALPLTSVVEVPEELIPKSGTFFWMDQTAFCRALLDIAQNTALR  
YKPSRLVLIODPFTTISQSRSE"  
5642..6406  
/gene="CPj0525"  
5642..6406  
/gene="CPj0525"  
/codon\_start=1  
/transl\_table=11  
/product="CT398 hypothetical protein"  
/protein\_id="BAA98731.1"  
/db\_xref="GI:8978896"  
/translation="MHDAISLILAEIDIMIRLMRYKKEHOKELAKVLSKSDIR  
KVOEKELEMENIKQIRIDGENRIOEISQIKLEMOQAAVKKMEFNALITQEMTANK  
ERRSLEHSLDMKQAGEDLIYSIKSLASTENSSVIEKELEFESIKRISGKAL  
LEQRTLEKHATNPESLSTYERLNKKRKYVPIENRYCSCGHYITLQHNELVAKKD  
RLIFCHESRLIYQESOVNAQENSTAKRRRAAV"  
/complement(6827..7816)  
/gene="kpsf"  
/complement(6827..7816)  
/gene="kpsf"  
/codon\_start=1  
/transl\_table=11  
/product="GutQ/Kpsf family sugar-P isomerase"  
/protein\_id="BAA98732.1"  
/db\_xref="GI:8978897"  
/translation="MPSPMISTDWCODILKOKAVDFEPOFAPQPKEMAOIAKTLIG  
SGWPFSSVCKSGCYARKLVATLDSLSRALFSPVDLHDLGVSGDLYCJFSKS  
GETQELDVTPEHLKSRALIVAITSMYNSIAALSDVLVILPVAEIDPFLIPNST  
TCQMIQDFDLMLFLHSGVSLSTYKMHKPGGVKMGVVKQFMPKPTVPCPHLG  
DKVSFSLVFSAYCGCYIVDPOFLMGITDGLRSLASGGEVLSLEKYMVA  
NPRCTTDSIDIALIQLEMSSPAVLPLVDNEENRHVTGLHMHLLAKAGLL"  
/complement(7843..9024)  
/gene="sucb\_2"  
/complement(7843..9024)  
/gene="sucb\_2"  
/codon\_start=1  
/transl\_table=11  
/product="dihydrolipoamide succinyltransferase"  
/protein\_id="BAA98733.1"  
/db\_xref="GI:8978898"  
/translation="MIEFRRPKIGETSSGSIYRMKLNLGDHVARDEPLIEVSTDKI  
ATELPSKAGRLVRECVNEGDEVASGDVLGILELEISEADESTSCGLTCEKRSKA  
GSSSSYMEPAVLSIAOREGIGLDNIOKTAGTCGRVRODLEAYISEQOYSTIE  
IFQGEVNRIPMSPLRRAIASLSKSDVEPHASLVVDVDTDLNMLISGERFLDTH

gene  
CDS  
complement(9038..10282)  
/gene="gltT"  
complement(9038..10282)  
/gene="gltT"  
/codon\_start=1  
/transl\_table=11  
/product="glutamate symport"  
/protein\_id="BAA98734.1"  
/db\_xref="GI:8978899"  
/translation="MKLMKIFIGLVGVTLGLVLEDKAIFPKPIGDIIFNLISMVY  
PLVCSNIGIASISDMKLGRIKISVGLIGTALIVIGLCFAMIFSPGNCDFP  
QAOQSDSAVTVDSNKTAYFLSTIAQVSPNVPKSPAEAGLIIQIIFALIGIALRL  
SGERGPRYERIDPESFIMLRMVNMLNFAFYQASWAMISGNHGLCVLMQAKRTI  
AYVLACLFHATLVFGLVRFCKKSFSEFLSSMDALSCAVSTASSSTLLVYATF  
KNLGSVAEVSFVPLGATVMNNGTALEQGAFAVIAQVNCPLSLSLLLVYATF  
SAVGSAGVPGGMITLGSVLASVGLPIQGIATLIGIDRLDVIQTPMNLGDVAVATY  
VASGEGELSPYESIKOESVET"  
complement(10282..11379)  
/gene="ycbH"  
complement(10282..11379)  
/gene="ycbH"  
/codon\_start=1  
/transl\_table=11  
/product="ATPase"  
Query Match 18.4% Score 301.4; DB 1; Length 300550;  
Best Local Similarity 50.3%; Pred. No. 1.3e-61;  
Matches 814; Conservative 0; Mismatches 776; Indels 27; Gaps 2;  
QY 2 AATATAAAACATATGATATGATAAATAAAGTATTTGAGAGGTAAATATGACAAAAC 61  
DB 106572 AACTAAATGCTATTAAGTCTTTATATAGCCATATAAGTTGGAGAGGTTATGCAATCATC 106513  
QY 62 CGAAGAAAAACCTTTTGGAAAAATTCGCTCTTTGTTGGCCGATACATCTACAGACT 121  
DB 106512 ACAGATGAACCCCTTTTCAAGGCTCGGGCATATCTTGCTCTATTTAAATCAGAAAT 106453  
QY 122 AAGAAAGTTCGCAATGTTCTTAATGTTCTGTAATTAATTAATTAATTAATTAAT 181  
DB 106452 TTCTAAGTTTGTCCACATTTCTACTACGCTTTTTCGTTGCTTAATTAATTAATTAAT 106393  
QY 182 ACGGATACAAAGACACTCTTATTTGAGAGCTCCTGTTCTGTTGCGAGAGCAATAC 241  
DB 106392 GAAAAACATGAAGATACCTGCTGATTTGCTGATCAGATGCTGGCGAGAGATTC 106333  
QY 242 TTTCAATCAAGTTTGGCTTTGTCCTGTCGTAATTAATTTATTTATTTATGCAAA 301  
DB 106332 CTTCCTTAAGGTTTGGGAATTTGTCGGAGACTTTATTTACTATGTTATGTTAGG 106273  
QY 302 GCTAGTATATTTTAAAGACAGGCTTATTTATGACAGTGGGAGAGCCCTTTTAA 361  
DB 106272 GTTAAAGAGTGGTATTCCTGGGATACGTTTATTTATTTATTTATTTATTTATTT 106213  
QY 362 TTTCTTGGCCGTGTTCCGACGTGTAATTTATCCGATAGCGATGTTTATTAATCTTACAGA 421  
DB 106212 TTTTCTTCCGTTTGTGCTGATCATTTATTCCTTAGGGAATACCCGATCTCAATC 106153  
QY 422 ATTTCGACCGTTTACAGCCATCTATCCAGAGATTTGCTAGAGATTCGTTGCCATCT 481  
DB 106152 TCTCGCTGATAAATTACAGAGCTCTTCCTCAAGAGCTTGCTGTTATTTATTTGATG 106093  
QY 482 AAGAACTGACATTTGCTGATTTATGTAATTTATGTAATTTATGTAATTTATGTAATTT 541  
DB 106092 CCGTTACGAGTTTACAGATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 106033  
QY 542 ATCTATATGTTCTGCGGATTTGCTAATGAATTTCAAAAATTCACAGAGCAAGACGTTT 601  
DB 106032 TTGATGTTTGTTCGGAGCTAGCCAAATCAATTTACTGAAGCGGCGGCTTT 105973

[illegible]

Accession	Version	Keywords	Source	Organism
M28816	GI:152469	ATP transport: ATP/ADP translocase.	R. prowazekii (strain Madrid E) DNA.	Rickettsia prowazekii
REFERENCE				Bacteria: Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; Rickettsiidae; Rickettsia; typhus group.
AUTHORS				1 (bases 1 to 1604)
				Williamson, L.R., Plano, G.V., Winkler, H.H., Krause, D.C. and Wood, D.O.
JOURNAL				Nucleotide-sequence of the Rickettsia prowazekii ATP/ADP translocase-encoding gene
MEDLINE				Gene 80 (2), 269-278 (1989)
PubMed				90060776
COMMENT				2555259
				Author's copy of sequence [1] kindly provided by D.O.Wood, 20-OCT-1989.
FEATURES	Source	Location/Qualifiers		
		1..1604	/organism="Rickettsia prowazekii"	
		/db_xref="taxon:782"		
		21..26	/note="-35_signal"	
		44..50	/note="-10_signal"	
		81..1577	/note="ATP/ADP translocase"	
		/codon_start=1		
		/transl_table=1		
		/protein_id="AAA26382.1"		
		/db_xref="GI:152470"		
		/translation="MSYKSSENVLSLELRKIIMPIEQYENKRFLEPLAMFICLLNYS		
		LRSLKDFGVVTDIGTESISFLKTYIVLPASVIMIIIVKLCIDILQCNVIVITSFL		
		GLFPLFAFVLEVPDYLPHDKITLESFLAVPNKFRKTIIGKGSFASFTVIELDCT		
		MMSLFMQFANOITKIAEAKRKYMSGILANILAPTSVIGYGLFHEKROIYAEHLK		
		FVPLFVIMITSSEFLIITLIRKMNVLDELRLYLPALYKREKTRKASTFESLKNITF		
		SKYGLALTLITAVGSVNLVEGWSKRVLELPKRAVITYMGOFOFQGWALIAM		
		LIGIONLIRKSVLMAAITPLMTITGAFAEFSFIFPQSVLMIATGLASPLTAVM		
		IGMIONLISGVKYSILPDATKMMVIFLDKDLRYKGOAAVEVIGRLGKSGALIOST		
		FFILPEVGFLEATPEPVASIFPIFVILIMIFAVKLNKEOVLVNKNK"		
BASE COUNT	537 a	201 c	257 g	609 t
ORIGIN				
Query Match	17.9%	Score 293.2	DB 1	Length 1604
Best Local Similarity	53.1%	Pred. No. 1.9e-59		
Matches 781	Conservative 0	Mismatches 653	Indels 36	Gaps 6
QY	66	GA AAAACCTTTTGGAAATGCGCTCTTTCTTGTGCGCCGATACATACAGAGCTAAAG	125	
DB	99	GAAATTTATCTTTGAGAACTAAGAAAGATAATTTGGCCTATAGAACAAATAGAAATAG	158	
QY	128	AAAGTTTGCACATCTTCCATATGTTCTTCGTATTACATTTAACTATAGCGTGTACGC	185	
DB	159	AGTTTGTGCCACTTCGATTTATGATGTTCTGTATTTATTAACCTACCACTTGTCTG	218	
QY	186	CATCAAAAAGACACCTTATTGTGTGGAGCGCTCGTGTCTGTGTCAGAGCAATACCTTTC	245	
DB	219	TCAATTAAGACAGTTTGTGTAGTACAGATATAGCTA-----CAGAAATCATAGTTT	272	
QY	246	ATCAAGTTTGGCTTGTGTGCTCCCTGCTGCTATTATCTTTATGCTTTATTAAGCAAGCTA	305	
DB	273	TTAAAAACATATATAGTACTACCTCTGCTGAATATGCTATGATTAATTAATGTTAAAGCTA	332	
QY	306	AGTATATTTTATAGTAAGACGCGCTATTATTATGAGAGGAGACGCCCTTTTATATTTC	365	
DB	333	TTTGATTTTAAACAGAAACGATATTTTATTTATTTACTCTCTTTTGTAGGGTAT	392	
QY	366	TTTGCCCTGTGCCAGCTATTTATTCGCGTACGAGATGTTTATCATCTTACAGAAATTT	425	
DB	393	TTTGCAATTTATGCTCTTGTCTTTATACCATATCTGATTAATTAAGTACACCGCATATAA	452	
QY	426	GCTGACCGTTTACAGGCACTCTACCTCCAGAGATTGTTAGAGATCTGTTCCATCTTAAGA	485	







AFSVSIFQNPQILLDEVEAAGDSYFEKSLNLMKNKFNTPISIVSHOEELIKDNC  
 DRCLIKGDHIIIDGTPSEIFRIYKQOSNKEIHK"  
 /product="(3R)-HYDROXYMYRISTOYL-[ACYL CARRIER PROTEIN]  
 /transl\_table=11  
 /db\_xref="GI:3860576"  
 /protein\_id="CA14481.1"  
 /db\_xref="GI:3860580"  
 /translation="MTAITETIMDLIPRYPFLVDRLKIDPNKSLIGIKNVNPE  
 OPTCHFPARPMVPGVIAWESMAOLAAILIIVAKSIDSTKKEVFLMSIENTKRRRIYQPG  
 DTMHHSYIDQORAWMKFSSKVMYECIEAESKRTAIIKDKM"  
 complement(7659. .8699)  
 /gene="Rp009"  
 /complement(7659. .8699)  
 /gene="Rp009"  
 /codon\_start=1  
 /product="UDP-3-O-(3-HYDROXYMYRISTOYL) GLUCOSAMINE  
 N-ACYLTRANSFERASE (1pxd)"  
 /protein\_id="CA14482.1"  
 /db\_xref="GI:3860581"  
 /translation="WVSNFYKNIGRPRLTAIVDFLHDIIETPKYEDIDYDIKILO  
 EASPDISFLSPKISEFLKTTKAACVPRNFTBEVNONTYLLHAENSYRASKLID  
 FFYAPTKSYSTRIMKSAIIADSATIGKNCYIGHNVIEDDVIIGDNLIDAGFTIGRG  
 VNIGKNARIQOHVSINVAIIGDDVILVGARIGDGFSETEKGVNHRKIFHIGIVKIG  
 NNVEIGSNMTTIDRCALODTIIEDLCRIDNLVQIGHGVIGKSGIIVAOAGIAGSSAIG  
 KYCALGGVGIAGHLNIGDGTQVAOGGVAQNIIEBCKIVGSPAPVIMDMHROSITMK  
 OLVKTSNKKTKK"  
 complement(9267. .9339)  
 /gene="tRNA Phe (GAA)"  
 /complement(9267. .9339)  
 /gene="tRNA Phe (GAA)"  
 9661.10666  
 /gene="Rp011"  
 9661.10666  
 /gene="Rp011"  
 /codon\_start=1  
 /product="NIFR3-LIKE PROTEIN (n1fr3)"  
 /protein\_id="CA14483.1"  
 /db\_xref="GI:3860582"  
 /translation="WLVNDIMLNPITFIKIGNIELSSNVILAPMSVDLDEFKRLV  
 KRFGAGLVNEMIAKRAMIMKRSQSMQSCAIIHMDPTACVOLACEDVDAEAKNM  
 EDMGAKIIDLNEGCPARKVGVGAGSALMBDEQLAKIFEATVKAVKIPVYKMKRIGN  
 DONTNAPETFLAKIANGSGVAVTVYGRRCOPYSGNAMDFRTYKBAVKIPIVLANGD  
 ITNFKAEALIDORSAGDIIIMGRGVYGGKDWLISOIAYLTKTGKKEKPARSIAEODIIT  
 KHYDAIIDYKGSVSPYARKHIIIMYSNGLPSSAFCRTVMIMKDPYAVKEKIAEFTM  
 SVMDBANK"  
 10683.11756  
 /gene="Rp012"  
 10683.11756  
 /gene="Rp012"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CA14484.1"  
 /db\_xref="GI:3860583"  
 /translation="MKRILSFEFLIFPNSSAYAKAIIIDYKPIFPIVUTENKKIK  
 IAIRSYLKNKRSYFLVDPNSKEETIVLOELVILPANKIEENLILKLNKSYIAALN  
 KYNFDIOSLSKODEFSNKKFTDNTIIVEHKLNSKSSHSSTFMNSTVOANYGANSMSY

gene  
 CDS  
 /gene="Rp006"  
 3757.5877  
 /gene="Rp006"  
 3757.5877  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CA14479.1"  
 /db\_xref="GI:3860578"  
 /translation="WVAATGHRICNDCKSAICYOKODEPVNIPIEBSNILEETLAPYG  
 LEIYILLTRMPLNIYAPLPEKPTNYNIIIVGLGAPAGSLASYLLRSHGNVTAIDGLK  
 ITPLSFNHMKPVKEWHEKKNLSEIRIPKGFQVAEYGITIRMDKNNLIDILIRERN  
 NKYYDVALGPNITRKBOALDLSFDHVAFCJGAGCKPKIINTEPEAKGVRTASDPLMT  
 LONGAFIQNSNTNMIMRPYIITIGGITSIDVATESLPTYYKQVEEPAKNTIEKDLT  
 EEDKIAIEEFLAHAKLEKAKNNEELKVFENKLGATVYRGRLODSPAYKUNHEBLI  
 YTLAIVNPKENMOPLRINIDKYGHVESVESITTMALDRTHKKYHIMKCPVFPSSN  
 TVLTKTIVMAIGIENNTQEDYDYSGDCNPKYFGSVYKAITSAKEGEYEVIRKL  
 IDNVPSFGSYACFTLODYLILSRINKINILNDKTEFLIHSPLAKNFOGOEFLR  
 ONYSKDIKLIIEPVALSPADIDIEKGLISFTIYEVYKSTSLCKTISENKKVILMGPTG  
 SPLEIPONKKITIIIDSKYRNGLKLIKLEKNKRVIFATYPPIDIKKKRLTSDVIVIIINTS  
 PELAELOELKIFGENTELLINVNSLMOCMMKNGICGCIQKVKOKYIFACSEQON  
 VEIIDFKSLKRLRONSLOEKMSN"  
 complement(6227. .7021)  
 /gene="Rp007"  
 complement(6227. .7021)  
 /gene="Rp007"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ACYL-[ACYL-CARRIER-PROTEIN]-UDP-N-  
 ACETYLGUCOSAMINE O-ACYLTRANSFERASE (1pxa)"  
 /protein\_id="CA14480.1"  
 /db\_xref="GI:3860579"  
 /translation="WNSNIHTTAIIAEGANFGKNVQPYCIIGPEVVLHDNVELKS  
 HVAIDGITEIGENTVITYPFSIGPPQILKANENSSYIIGSNMTIRERYVVOAGSKS  
 GGMITRVGNNNLFFMVGHIGHCKIGNNLVFANYSLAGHAKVGDYAIIGGISAIVHOY  
 TRIGEYSGISPGVADYIPGVLSSRAVEYENGLNMGMRKGGKADSLTALNAVE  
 EFLIGEGNPDVRIKOVAREKKNNSIVTQIIDLNDSSRACHFEKK"  
 complement(7031. .7468)  
 /gene="Rp008"  
 complement(7031. .7468)  
 /gene="Rp008"

gene  
 CDS

Query Match 17.9% Score 293.2; DB 1; Length 282610;  
 Best Local Similarity 53.1%; Pred. No. 1.2e-59;  
 Matches 781; Conservative 0; Mismatches 653; Indels 36; Gaps 6;

gene  
 CDS  
 /gene="Rp012"  
 10683.11756  
 /gene="Rp012"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CA14484.1"  
 /db\_xref="GI:3860583"  
 /translation="MKRILSFEFLIFPNSSAYAKAIIIDYKPIFPIVUTENKKIK  
 IAIRSYLKNKRSYFLVDPNSKEETIVLOELVILPANKIEENLILKLNKSYIAALN  
 KYNFDIOSLSKODEFSNKKFTDNTIIVEHKLNSKSSHSSTFMNSTVOANYGANSMSY

gene  
 CDS  
 /gene="Rp009"  
 complement(7659. .8699)  
 /gene="Rp009"  
 /codon\_start=1  
 /product="UDP-3-O-(3-HYDROXYMYRISTOYL) GLUCOSAMINE  
 N-ACYLTRANSFERASE (1pxd)"  
 /protein\_id="CA14482.1"  
 /db\_xref="GI:3860581"  
 /translation="WVSNFYKNIGRPRLTAIVDFLHDIIETPKYEDIDYDIKILO  
 EASPDISFLSPKISEFLKTTKAACVPRNFTBEVNONTYLLHAENSYRASKLID  
 FFYAPTKSYSTRIMKSAIIADSATIGKNCYIGHNVIEDDVIIGDNLIDAGFTIGRG  
 VNIGKNARIQOHVSINVAIIGDDVILVGARIGDGFSETEKGVNHRKIFHIGIVKIG  
 NNVEIGSNMTTIDRCALODTIIEDLCRIDNLVQIGHGVIGKSGIIVAOAGIAGSSAIG  
 KYCALGGVGIAGHLNIGDGTQVAOGGVAQNIIEBCKIVGSPAPVIMDMHROSITMK  
 OLVKTSNKKTKK"  
 complement(9267. .9339)  
 /gene="tRNA Phe (GAA)"  
 /complement(9267. .9339)  
 /gene="tRNA Phe (GAA)"  
 9661.10666  
 /gene="Rp011"  
 9661.10666  
 /gene="Rp011"  
 /codon\_start=1  
 /product="NIFR3-LIKE PROTEIN (n1fr3)"  
 /protein\_id="CA14483.1"  
 /db\_xref="GI:3860582"  
 /translation="WLVNDIMLNPITFIKIGNIELSSNVILAPMSVDLDEFKRLV  
 KRFGAGLVNEMIAKRAMIMKRSQSMQSCAIIHMDPTACVOLACEDVDAEAKNM  
 EDMGAKIIDLNEGCPARKVGVGAGSALMBDEQLAKIFEATVKAVKIPVYKMKRIGN  
 DONTNAPETFLAKIANGSGVAVTVYGRRCOPYSGNAMDFRTYKBAVKIPIVLANGD  
 ITNFKAEALIDORSAGDIIIMGRGVYGGKDWLISOIAYLTKTGKKEKPARSIAEODIIT  
 KHYDAIIDYKGSVSPYARKHIIIMYSNGLPSSAFCRTVMIMKDPYAVKEKIAEFTM  
 SVMDBANK"  
 10683.11756  
 /gene="Rp012"  
 10683.11756  
 /gene="Rp012"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CA14484.1"  
 /db\_xref="GI:3860583"  
 /translation="MKRILSFEFLIFPNSSAYAKAIIIDYKPIFPIVUTENKKIK  
 IAIRSYLKNKRSYFLVDPNSKEETIVLOELVILPANKIEENLILKLNKSYIAALN  
 KYNFDIOSLSKODEFSNKKFTDNTIIVEHKLNSKSSHSSTFMNSTVOANYGANSMSY

gene  
 CDS  
 /gene="Rp008"  
 complement(7031. .7468)  
 /gene="Rp008"

gene  
 CDS  
 /gene="Rp006"  
 3757.5877  
 /gene="Rp006"  
 3757.5877  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="CA14479.1"  
 /db\_xref="GI:3860578"  
 /translation="WVAATGHRICNDCKSAICYOKODEPVNIPIEBSNILEETLAPYG  
 LEIYILLTRMPLNIYAPLPEKPTNYNIIIVGLGAPAGSLASYLLRSHGNVTAIDGLK  
 ITPLSFNHMKPVKEWHEKKNLSEIRIPKGFQVAEYGITIRMDKNNLIDILIRERN  
 NKYYDVALGPNITRKBOALDLSFDHVAFCJGAGCKPKIINTEPEAKGVRTASDPLMT  
 LONGAFIQNSNTNMIMRPYIITIGGITSIDVATESLPTYYKQVEEPAKNTIEKDLT  
 EEDKIAIEEFLAHAKLEKAKNNEELKVFENKLGATVYRGRLODSPAYKUNHEBLI  
 YTLAIVNPKENMOPLRINIDKYGHVESVESITTMALDRTHKKYHIMKCPVFPSSN  
 TVLTKTIVMAIGIENNTQEDYDYSGDCNPKYFGSVYKAITSAKEGEYEVIRKL  
 IDNVPSFGSYACFTLODYLILSRINKINILNDKTEFLIHSPLAKNFOGOEFLR  
 ONYSKDIKLIIEPVALSPADIDIEKGLISFTIYEVYKSTSLCKTISENKKVILMGPTG  
 SPLEIPONKKITIIIDSKYRNGLKLIKLEKNKRVIFATYPPIDIKKKRLTSDVIVIIINTS  
 PELAELOELKIFGENTELLINVNSLMOCMMKNGICGCIQKVKOKYIFACSEQON  
 VEIIDFKSLKRLRONSLOEKMSN"  
 complement(6227. .7021)  
 /gene="Rp007"  
 complement(6227. .7021)  
 /gene="Rp007"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ACYL-[ACYL-CARRIER-PROTEIN]-UDP-N-  
 ACETYLGUCOSAMINE O-ACYLTRANSFERASE (1pxa)"  
 /protein\_id="CA14480.1"  
 /db\_xref="GI:3860579"  
 /translation="WNSNIHTTAIIAEGANFGKNVQPYCIIGPEVVLHDNVELKS  
 HVAIDGITEIGENTVITYPFSIGPPQILKANENSSYIIGSNMTIRERYVVOAGSKS  
 GGMITRVGNNNLFFMVGHIGHCKIGNNLVFANYSLAGHAKVGDYAIIGGISAIVHOY  
 TRIGEYSGISPGVADYIPGVLSSRAVEYENGLNMGMRKGGKADSLTALNAVE  
 EFLIGEGNPDVRIKOVAREKKNNSIVTQIIDLNDSSRACHFEKK"  
 complement(7031. .7468)  
 /gene="Rp008"  
 complement(7031. .7468)  
 /gene="Rp008"

gene  
 CDS

Db	59434	TCATTTAAAGACGGTTTGTGACTAACACATATAGGA-----CAGAAATGCATTAAGTTT	59487
Oy	246	ATCAGATTGGCTTGTGTGCCCTGTGCTATTATCTTATGCTTATTTATGCAAGCTA	305
Db	59488	TTTAAACATATATAGTACTACCTCTGCTGTAATTGCTATGATTAATATCTTAAAGCA	59547
Oy	306	AGTATATTTTAAAGACAGGCTTATTTTATGAGTGGGAACGCCCTTTTAAATTTTC	365
Db	59548	TGTGATTTTAAACAGAAAGCATTTTATGTTATTACTTCATTTTTTTAAAGGAT	59607
Oy	366	TTTTGCCCTGTCCGACATGTAAATTTATCCGTACAGCATGTTTATCCTACAGAAATT	425
Db	59608	TTTGCAATTATTTGCCCTTGTGTCTTACCACTATTCGATTTAGTCCACCCCTATCATAA	59667
Oy	426	GCTGACCGTTTACAGGCCATCCTACCTCAGAGATTGCTAGACCTGTGCCATCTTAAGA	485
Db	59668	ACTATAGATCTTTTAAAGTTTACCTTATCTTAAATTTTCAATGGTTTAAATAATGTGTT	59727
Oy	486	AACGTGACATTTTGGCATTTTATTTACTGCTGTGAACTATGGGGAAGGTCATGCTATCT	545
Db	59728	AAATGAGCTTTGCACTCTTTTATCTATTTGCCGAGCTTTGGGGAACATGATGCTTGGT	59787
Oy	546	CTAATGTTCTGGGGATTTGCTAATGAATTTACAAATTTCAAGAAAGCAAGCGTTTCTAC	605
Db	59788	TTATATTATTGGCAATTTGCTATCAATTTCAATTAATGCGTGAAGCTAAAGCTTCTAC	59847
Oy	606	GCTCTTTTCGGTATCGGACGTAAATTTTCTTACTAGCTTCTGTCTGCATTTGTTGG	665
Db	59848	TCAATGTTTGGTTTCTTGCATTTTACCATTCCTGTAACATCAGTGGTTATTTGATAT	59907
Oy	666	GCTTCAAGTTGAGAGCTTCGGTTTCTGAAGTGAATGCTTGGGAATTTCTTACGT	725
Db	59908	TTTCTACAGAAAAACT-----CAATATGTTGCGAACAATTAAAAATTTGTA	59955
Oy	726	CTTTTGATGGCATGACTATTGTATCTGGACTTGTCTTATGGCCAGTTACGTGATGTC	785
Db	59956	CCTTATTTGTTATATGATAAACAGTAGTTTCTTAATAATTTAACATATGATGAGATG	60015
Oy	786	AATAAGAACGTAATGACCGATCTCGCTTCTATTAATCCAGAAAGATGCAAAAGGGA	845
Db	60016	AATAAAATGTCTAACGATCTTACGACTATATGATGTCAGCATTTGATGTAATAA-----GAA	60069
Oy	846	AAAGTGCTAAACCTAAATGAATGAATGAAGTACCTTCCTATCTGATGATGATCTCT	905
Db	60070	AAAAAACTTAAAGCTAAATTTGTCGTTCAATAGAAATTTAAAAATGATCTTTACTTCGAA	60129
Oy	906	TATATTTCTTTAATTAACCTCTGTGGTATATTCGCTATGATTTTGCAATTAACATCGAA	965
Db	60130	TATGATGTTTATTTGCACTATTATTAATTTGCTTATGTTGTTTCGTAATTTTACTGAA	60189
Oy	966	GTGACTTGGAAAGTCAGCTGAACCTGCAATTCCTAATATGATGACTATATGAGTTC	1025
Db	60190	GGTGTGGGAATCCAAAGTAATTAATTAATTCGCAAAAGGAGGCTTAAACCATATAT	60249
Oy	1026	ATGGGAACCTCTCCTCTGACAGGCGCTAATTCGCACTATTCATCTATTTCTGCT	1085
Db	60250	ATGGGTGCTTCCATTTTATCAAGGTGGGTGCAATTTGCTTTTATGTC---TGATAGT	60306
Oy	1086	GGTAACGCATTCGTAATTTGATGATTTAACTGAGCCATAGTCACTCTGTCATGCTT	1145
Db	60307	AGTATATTTTAAAGAAAGTATCATGCTTAAGTACGATATGATCATCTCATTAATGATG	60366
Oy	1146	CTCTCAACAGGTATCGTTTCTGCGTCTGTATCTTTAGAAACCAAGCTTGGGCTG	1205
Db	60367	TTCAATTAACAGGTGGGCACTTTTTCATTTTATTTATTTTGTAGAGGTTATGCAATGAT	60426
Oy	1206	GT----CGCTATGTTGCGTACACATCTCTCATGCTAGCTGTGGTTGTGGAGCTATACAG	1262
Db	60427	TTTAAACGCACTCTTGTCTCAAGTCTTAAACACTGCTGTATGATGCGTATGATGTCAA	60486
Oy	1263	AATATTTCTTTGGAATTCACAAATTAAGCTCTGCTTTTGCTACATGAAGAAATGGCCAT	1322
Db	60487	AATGTTTAAAGTGAAGGTGTCAAATTTCTTTATTTGATGACACTAAATAATATGCGGAT	60546

OY	1323	ATCCCTTGAGACCAAGCAAAATGCAAGGTAAGCGTCAATGATGATGATGCGCC	1382
Db	60547	ATTCCACTTTGATAGAGATTTTACGAGTCAAAAGGCGACGCTCCGTGACGTTATCGAGGA	60606
OY	1383	CGCTTGGCAAAATCAGAGGAGCTTTAATCCACAAGCTTTTCTCGTTATTCGTGGAGCT	1442
OY	1443	ATTGAGAGCTAT-----GACCCCTATCTCGAGGATTCCTTTTCATCATTTGCTATT	1496
Db	60607	AGGCTCGGTAATTAAGCGCGCTCTATTATTCATCTACATCTTATTATTTATTCCTGTA	60666
OY	1497	TGGTTGTTTCTGCACTAAGTTAAACAA	1526
Db	60667	TTTGGTTTATAGAGCGACCTCTATTGCTGCTATATTTCTTATATATTAAGTAATATTA	60726
Db	60727	TGGATATTTGCAAGTTAAAGTTAAATAA	60756
RESULT 24			
LOCUS	AE008575	14117 bp	DNA
DEFINITION	Rickettsia conorii Malish 7, section 7 of 114 of the complete genome.		
ACCESSION	AE008575		
VERSION	AE008575.1		
KEYWORDS	GI:15619111		
SOURCE			
ORGANISM	Rickettsia conorii.		
REFERENCE	Bacteria: Proteobacteria; alpha subdivision: Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.		
AUTHORS	1 (sites)		
TITLE	Ogata,H., Audic,S., Barbe,V., Artiguenave,F., Fournier,P.E., Raoult,D. and Claverie,J.M.		
JOURNAL	Selfish DNA in protein-coding genes of Rickettsia		
REFERENCE	Science 290 (5490), 347-350 (2000)		
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE	Mechanisms of evolution in Rickettsia conorii and R. prowazekii		
JOURNAL	Science 293 (5537), 2093-2098 (2001)		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE	Direct Submission		
JOURNAL	Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 13385, France		
COMMENT	A public version of R. conorii genome database is accessible at <a href="http://lrs-server.cnr-mrs.fr/">http://lrs-server.cnr-mrs.fr/</a> . The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.		
FEATURES	location/Qualifiers		
SOURCE	1..14117		
gene	/organism="Rickettsia conorii"		
	/strain="Malish 7"		
	/db_xref="taxon:781"		
	127..1809		
	/gene="yidC"		
	/note="RC0074"		
	127..1809		
CDS	/gene="yidC"		
	/codon_start=-1		

```

/transl_table=1
/product="60 kD inner-membrane protein"
/protein_id="AAL02612.1"
/db_xref="GI:15619112"
/translation="MNNITLIALIILSLISIFGWOYFVVKPEQKKOQOIAVQKAE
NLKQQLKALVEPATGIYQVESQVQKIKIESESLTGISLKGRLPDDLKKKQDL
SKNSPEVRLFSPTANTENAYFAEVLGNLSLKVLPNDITWNSDSEILSEKPYHLW
VNEDEVFLVITVDENYLFTEOTIVNSDKELPYQSYGLINRKYIAVEKAVNLHQ
GPICIDENLKESYDDIDKDKSEKFAKSVDMGTIDKYWLSLIPDKSSNSNPN
YALKQCTERYQVDFISPVQITIKRGENSISKRIFGAKKVDLDKYEKQIDIKLFDA
IDGWFYITIKPEVYAMNFFGVGNFVSLITVYTIKLMFPLANKSYRSMKMN
LQPEIDRIKMLYSDDKARLNOEIMALKKKEVNVACPLVQIPVFFSLYKVLVT
IEMRAQPFYGIWKDLSASDPTIFGLLPFSPPSLMIGAMPILMAITMFLQOKKS
PEPADPMQAOVMKFMPLIFLMFSSFPVGLLIYMSNNIILIIQOYIYNKN"
1813..2358
/gene="pgsa"
/ncle="RC0075"
1813..2358
/gene="pgsa"
/codon_start=1
/transl_table=1
/product="CDP-diacylglycerol--glycerol-3-phosphate
3-phosphatidyltransferase [EC:2.7.8.5]"
/protein_id="AAL02613.1"
/db_xref="GI:15619113"
/translation="MRIDKNLPNYLFIARIWVPIITLVPIYNNSLARTIGALLFVLA
SIMPFGYIARKKNYLVTSGKMPDPIADKILVGCYIIMLKKNVDEIPCLILAB
FLVSGLEPLALVAVSVSRKAVKTFLOWFALSIILLSKSGSGLIYLDVGEITIM
IAFLTITGTSTYKACKRYF"
2372..2932
/gene="RC0076"
2372..2932
/gene="RC0076"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL02614.1"
/db_xref="GI:15619114"
/translation="MIHKYIIIGFIWSSNLKYAANNDIANKKYKVEVSEHKT
LNEASTLISCVDFRLIDETDKLKKQJGLEENFKVSLPGASLILNDKTYWKTE
DTEILLOELHMIKOIPELDHRECGAYTILIGEOELNKEKETAAHAAILNKARDIIE
KIPQLKYVTEFGDGVGEVQEIIPS"
complement(2964..3063)
/ncle="REP01, repeated element"
complement(3102..3521)
/gene="RC0077"
complement(3102..3521)
/gene="RC0077"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL02615.1"
/db_xref="GI:15619115"
/translation="MVLRSRLRLKVMYIKKIIIFGIALILSTSPANSTYSDGSKDA
KTNIDITOKIIDDSAYAGTITKPEVREIOKYRAIYKINKKRELKYNRLSKENQNL
AEQOKYKOKLSISKLIVENDOKNNTADSNKSKDTR"
complement(3542..3685)
/ncle="REP05, Rickettsia palindromic element 1"
3705..3977
/gene="RC0078"
3705..3977
/gene="RC0078"
/codon_start=1
/transl_table=1
/product="similarly to BioC"
/protein_id="AAL02616.1"
/db_xref="GI:15619116"
/translation="MYFLFVCYCLVSNILIVGKNIOYYNNAOEFSRTINADLSN
YKATSYLPQAHIVDAGCGVCDTKFELSORYVYTAFDSTENK"
4544..5392
/gene="RC0079"
4544..5392
/gene="RC0079"

gene
CDS

/transl_table=1
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAL02617.1"
/db_xref="GI:15619117"
/translation="MAYRTNAVAIAINVFEEIAGSDDRPGLTLMIDSKLFSRLQQL
SLIITIDGNQIOTFASAKILVIVJDKSVINQWYFNLNIDIFYNNVMASTLTFST
NKELLIDGNISKIPVTAISKHGNQDIPQKGFILSPQATSLPVVNIINDSVKLIIEFD
KDKGLINLSTASTIYTGPIPLVQNGKVVNDPKDDAHARTALGVCNDCTIYIYVE
HITKQHVADKLTVQVRSILRQEKIINDKLIIPALKILKHLVNTVIGLTYTELAD
YMLTWI"
complement(5609..6283)
/gene="def3"
/ncle="RC0080"
complement(5609..6283)
/gene="def3"
/codon_start=1
/transl_table=1
/product="polypeptide deformylase [EC:3.5.1.31]"
/protein_id="AAL02618.1"
/db_xref="GI:15619118"
/translation="MSKLTALALITAIIFITGCKSNMKNENNISLPQYVTLNIQTLS
SEDKTIRIKAKTLNPLSEDLRDISILKRYQDEENCGALAAQOISGCIITFAVH
EADLKKMHPDLKDTMKTITWINSYKPIGIDKHEDYEGCFSVNATGPARFKKIY
HAYDINGNQJQIAGFPLAVIQHEIDLHNGKVLQDVAPKKIKMTKEYLEMKKAME
QENIKS"
6482..7978
/gene="tlc1"
/ncle="RC0081"
6482..7978
/gene="tlc1"
/codon_start=1
/transl_table=1
/product="Adp, Arp carrier protein"
/protein_id="AAL02619.1"
/db_xref="GI:15619119"
/translation="MNNPKNDNYLSELSKYIMPIERYENKFKFLPMAPMFCILLNST
LRSTKDGFWVTIDGAEALISPLKTYIVPSAVIAMYIKICDILKONVYVTEFSPL
GYALFELFVLYPRDVLVHPDETIESNVAVPNKKPTRYGKKSFRSTYMAELK
MMLSLTFWQFANQITTKDEAKRFYSMGLANLALPVTSVITGCLHEKSTFYIAELHK
FVPLVAMITISSFLVILTYRMKNKNVLTDLPLDPAVKKKAKAKNSLIDSEFMIFT
SKYGVYATALLIAYGVSVNVEGVAKSVRELYTKEDAYTIYMGKFPQYGVVAIAM
LIGSNIILRKYSWILFAAMITPLMMIITGAEPFAIFPDSVIAHLTGILASGPLAVM
IGMTQNLVSGVKYSLFDPATKMAVITPLDDKIDRYKGAAYEVJGGRGCKSGALIOST
FFILFPAFGVEATPYFASIFEVYIIMTYAVKLNKEYVLVAKTEK"
8175..9470
/gene="g1pt"
/ncle="RC0082"
8175..9470
/gene="g1pt"
/codon_start=1
/transl_table=1
/product="glycerol-3-phosphate transporter"
/protein_id="AAL02620.1"
/db_xref="GI:15619120"
/translation="MWTNRSKITPRSPYKKVYKNSMVRILYSIIIGYATPYFCRQNF
NIATPALRETFGVTKIQIGWLTASSIMYGVKACNGFISDKVNARLFWVLGLFVGI
ITMLIGFSDSLMILGILMIASNMFQSGMWEPATKMLTHWFAKSKELGKMGATSNQI
GGALAMISCGYLIDKPFMRAAFVPGVACVSLFNLHWRNSKPEYGLSTVEYKXY
PLEAIGYERKLIPQLKAVCNKLIYVCANNEYIIRSGVYMAFPLKDLRNTS
LANAGLIGIYEMGIRGALIAVLSLRKLOGRNGFPVASTICMVLISILFMPKLPFH
SELSTYIILSLIGFFVSGPOLVAGIAADSTRAVGTANGLSGLFGLCAIAGVAV
GWISDNYGMNGVLFETISALLGGLPALIMNSAKRYI"
9495..9917
/gene="ndk"
/ncle="RC0083"
9495..9917
/gene="ndk"
/codon_start=1
/transl_table=1
/product="nucleoside diphosphate kinase [EC:2.7.4.6]"
/db_xref="GI:15619121"

```

/translation="MTIOYTFSMKPKDAIKRNKIGQVNTYIENAGLKIIVAKMKFLRK

Query Match 17.7%; Score 289.4; DB 1; Length 14117;

Best Local Similarity 52.8%; Pred. No. 1.3e-58;

Matches 808; Conservative 0; Mismatches 681; Indels 40; Gaps 7;

QY 7 AAAACATCAGAAATAGAAATATTAAGATTTTCAGAGGGTAAATATGACAAAACCGAG 66  
 Db 6445 AATACCTTATCTTCAAAAATAAGAAAATTTAATTTATGAATATCCCAAAA-----ATG 6500  
 QY 67 AAAACCTTTTGGAATAATGGCTCTTCTTGTCGCCGATACATPACAGAGCTAAAGA 126  
 Db 6501 ACAATTATCTCCGAACTAGCAAGGTAATTTGGCTTATGAAGAATATGAATAATAGA 6560  
 QY 127 AAGTCTGCAATGTTCTCAATGTTCTCTGATTATACATTTAACTATACGGTGTACGG 186  
 Db 6561 AGTTCTCTCATGCGCTTATATGATGCTTCTGTATTTTATTAACACTCAACGCGTCTGTT 6620  
 QY 187 ATCAAAAGACACTTATATGCGAGCTCCGCTGCTGGTGCAGAGCAATACCTTTCA 246  
 Db 6621 CAATTAAAGACGGGTTTGTAGTAA-----CGGATATAGAGGACAGAACCAATAGTTTCT 6674  
 QY 247 TCAAGTTTGGCTTGTGCTTCCCTGCTATATCTTATGCTTATTTATGCAAGCTAA 306  
 Db 6675 TAAACATATATATAGTACTACTTCTGCGCTATGCGCTATGCTATGTTATTTATGCTCT 6734  
 QY 307 GTATATTTTAAAGTAGAGGCGCTTATTTATGACAGTGGAAACGCCCTTTTAAATTTCT 366  
 Db 6735 GCGATATTTTAAAGCAAGAAAACGCTTTTATGTTATTTACTTATTTCTTAGGGTANT 6794  
 QY 367 TTGCCCCCTTCCGACCTAATTTATCCGCTACGCGATGTTTACATCTCCACAGATTTC 426  
 Db 6795 TTGCTATTTATTTGCGTTTGTCTTCTTACCCCTTACCCGATTTAGTTGCTCATCTGAAA 6854  
 QY 427 CTGACCGTTTACAGGCCATCTACCTCCAGATGCTTACGACTGCTCCATCTTAAGAA 486  
 Db 6855 CCATAGATCTTGGAGTGTACTTCTTCTATGTCATATGCTTATTAAGATATGTTGGA 6914  
 QY 487 ACTGACATTTTCTGCATTTTATGTTACTTGTGAACATATGGGGAACGTCATCTATCTC 546  
 Db 6915 AATGAGTTTTCATCTTTTATATCTATGAGAGAGCTTTGGGAAACATATGCTTAGTT 6974  
 QY 547 TAAATGCTGGGANTTCTGCTATGAAATTAACAAAATCCAGGAAGCAAGGCTTTCTACG 606  
 Db 6975 TATATTTTGGCAATTTGCTAACCAATTAACCAACGATGAGCTTAAACGTTTACT 7034  
 QY 607 CTCTTTTCTGATCGAGACTAATATTTCTTACTAGTCTTGTGCTGCAATTTGTTGG 666  
 Db 7035 CAATGTTGGTTTACTGCTATTTAGCATGCGCTTAACATCCGTAATTTTGTGATTT 7094  
 QY 667 CTTCAAAGTTAGAGCTTCCGTTTCTGAAGGTGATGATCTTTGGGAATTTCTTTACGTC 726  
 Db 7095 GCTTCACGAAAAAACT-----CAAAATGTTGACAGAACATCTAAATTTTGTGC 7142  
 QY 727 TTTTGATGGCTATGCTATGCTATCTGAGCTTGTCTTATGCGCAGTACTAGGAGATCA 786  
 Db 7143 CTTTATTTGTTATATGATTAACAAGTACTTTTGGTATATTTAACATATAGATGAGATGA 7202  
 QY 787 ATAAAGAGTATGACCGATCTCGCTTCTTATTAATCCAGAAAGATGCAAAAGGAGAAA 846  
 Db 7203 ATAAAGAGTCTTAACGATCTCTAGACTTTATGATCCGGCATTTAGTAAACAAAAGAAA- 7261  
 QY 847 AAGTGCTAAACCTTAATAATGAATATGAAGATGCTTCTTATCTGTGATGATCTCTCT 906  
 Db 7262 -----GCTAAAGCTAAATATGATTAATAGACACTTTTAAATGATCTTTACTTAAGT 7316  
 QY 907 ATATCTTTTATTAACCTCTTGGTATATGCTTATGCTATGATTTGATTAATCTAATCGAG 966  
 Db 7317 ATGATGTTATATGCTATGTTACTATGCTTACGGGTTTCAGTAAATTTAGTTGAG 7376  
 QY 967 TGACTTGAAAAAGTCCAGCTGAACCTGCAATATCTTAATATGAATGATAGTCACTTCA 1026  
 Db 7377 GTGTTTGGAAATCCAAAGTAAAGATTTATATCGACGAAGAGGCTTACACTATATATA 7436

QY 1027 TGGGAACTCTCTCTTCGACCTGGCTAGTATCGTACTTATCATGCTATTTGTTGGTG 1086  
 Db 7437 TGGGAAGTCCAAATTTATCATAGGCTTGGCTTCCAAATTCGTTTCATGC---TTATAGSTA 7493  
 QY 1087 GTACGTCATTCGTAATTTTGGATGCTTAACTGGAGCCCTTGTACCTCTCTCATGTTC 1146  
 Db 7494 GTAAATTTTAAAGAAAGTATCATGCTTAACTGCAAGCTATGATATCTCATTAATGATGT 7553  
 QY 1147 TCCTAACAGTATGCTTCTTCTGCTGCTTGTATCTTATAGAACCAAGCTCTGGGCTGG 1206  
 Db 7554 TAATTACCGGTGACAGCGCTTTTTCGATTTATCTTTTGTGATGTTATTTGCTATTCGATT 7613  
 QY 1207 TCGC---TATGTTGGTACAACTCCTCTCAATGCTAGCTGTGTTGTCGACCTTATACGA 1263  
 Db 7614 TAACAGGATATCTGCTTCAGGCTCCTTACGACTGCTGTTATGATGCGTATGATTCAAA 7673  
 QY 1264 ATATTTCTTGCAAATCCCAAAATATAGCTCTCTTTGACTCAACTAAAGAAATGGCTATA 1323  
 Db 7674 ATGTTTAAAGTAAAGGTGTAAATATCTTTATTTGATGCGCACATAAAACATATGGCGTATA 7733  
 QY 1324 TCCCTCTTGACCAAGACGAAAGTCAAAAGGTAAGGCTGCTATGATGATGTTGCCGCC 1383  
 Db 7734 TTCCGCTTGATAGAGATTTTACGATTAAGAGGCAAGCTGCTGTTGAAGTTATCGAGAGAA 7793  
 QY 1384 GCTTCGAAAAATCAGAGAGAGCTTTAATCCAACAAGGTTTGCCTGTATCTGT-----G 1437  
 Db 7794 GATTTGGTAATCAGAGGCTGCTATTTATTCATATCTTATTTATTTATTTTCCTGCAT 7853  
 QY 1438 GAATATTTGGAGCTATGACCCCTTATCTTGCAGATGATTTCTTTTATCATCTTCTATT 1497  
 Db 7854 TTGGTTTGTAGAGCAACCTCTTATTTTGTCTTCTATATTTCTTGTATATGTAATATTAT 7913  
 QY 1498 GCTTGTTTGTGCAACTAAGTTAAACAAA 1526  
 Db 7914 GGATATATGCCGTTAAAGGTTAAATATAA 7942

RESULT 25  
 LOCUS A.tAATP2 2139 bp mRNA Linear PLN 29-APR-1998  
 DEFINITION A.thaliana mRNA for AATP2.  
 ACCESSION X94626.1 GI:1707363  
 VERSION X94626.1 GI:1707363  
 KEYWORDS AATP2 protein; adenylate translocator.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;  
 Rosidae; eurosoids II; Brassicales; Brassicaceae; Arabidopsids.  
 REFERENCE 1 (bases 1 to 2139)  
 Mohlmann,T., Tjaden,J., Schwoppe,C., Winkler,H.H., Kampfenkel,K.  
 and Neuhaus,H.E.  
 Occurrence of two plastidic ATP/ADP transporters in Arabidopsis  
 thaliana L. --molecular characterisation and comparative structural  
 analysis of similar ATP/ADP translocators from plastids and  
 Ricketlesia prowazekii  
 Eur. J. Biochem. 252 (3), 353-359 (1998)  
 MEDLINE 98206726  
 PUBMED 9546649  
 REFERENCE 2 (bases 1 to 2139)  
 Neuhaus,E.  
 Direct Submision  
 Submitted (03-JAN-1996) E. Neuhaus, Universitaet Osnabrueck,  
 Pflanzenphysiologie, Barbarastr. 11, D- 49069 Osnabrueck, FRG  
 TITLE JOURNAL  
 JOURNAL

#### FEATURES

source

1..2139  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 82..1791  
 /function="adenylate translocator"  
 /codon\_start=1  
 /product="AATP2"

```

/protein_id="CA64329.1"
/db_xref="GI:1707364"
/translation="MEGLIOTRGLISLPASHRSKVLDPSSHGLKORLEFTTNLPALISLS
LMTVRNRPFSKSHLGRFPTRREHEDSLARKLRPRRKCVDESDTAAVSPRTIG
VEVTLTKRIVPLGMEFCILFENYTLRDTKDVLTAVRAGSSAEIIPELKTVNPMAL
GFMILYRLISMVSKKALFYVIVPFIYFAPGFVMPRNLQIPEALAKLLATLG
PRFMPLAIRMVSWFCLEFYVMAELMGSVYVLPFMFGVNPQITTYDEAKFPLFGLG
NALIFSGRTKVESSNMKNLPGVDGVAIVLKAMSVIYMGILACIETLMMVNRVYP
LPTRSKKKVYKPMQTMESLKFIVSPYIRLATLVAYGISINLVEVTKMSKLSOP
PSPDEYSKMGDFSTCTGATFTFMILSQQYVFKRGVWVAKIPPTVLLTGVAFSL
ILFGPEAPLVIKGMTPPLAAVYVPEVSVANVOHSSPTPMOCLPLDEVSK
VKAKIOLMMSATIKSGGALLQOFMILFGLANSTPYLGLIIVTAMLAASKLE
GPV"

```

BASE COUNT 540 a 456 c 503 g 640 t  
ORIGIN

Query Match 17.3%; Score 283.8; DB 8; Length 2139;  
Best Local Similarity 52.8%; Pred. No. 3.3e-57;  
Matches 749; Conservative 0; Mismatches 637; Indels 33; Gaps 5;

```

QY 120 CTAAGAAGTCTGCCAATGTTCTTAATGTTCTGTGATTAACATTTAAGCATG 179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 403 CTGAAGAAGATTGTCCTTAAGGCTTAATGTTCTTTCATCTTTTCAATTAACAATC 462
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 180 TTACGGATACAAAGACACTCTTAATGCGGAGCTCTGTCGTGTCAGAGGCAATA 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 463 CTTAGGACACAGAGATGTTTGGTGTGACGGCTAAAGAAAGTTCTGTGATTAATA 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 240 CCTTTCATCAAGTTTGGCTGTGTCCTGCTGCTATTAATCTTAATGCTTAATGCA 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 523 CCTTTTGAAGACATGTCGTAATGTCGATGCTATGAGTTATGTTGCTATACACC 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 300 AAGCTAAGTAATTTTAAGTAAGACAGGCTTATTTTATGACGTGGAAAGCCCTTTTA 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 583 AACTTTCCTCAATGTTCTCCAAAAGGCTCTTTTACAGCTTAATGTCCTCTTCAT 642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 360 ATTTCTTTCCTTCCCGACTGTAATTTATCCGCTACGGAGTTTATACATCTTACA 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 643 GCTACTTTTGGAGCCTTTGGTTCGTGATGACCTCCGACGAATTTATTAAGCTGAA 702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 420 GAATTTGCTGACCGTTTACAGGCATCTCACTCCAGAGTTGCTAGACTCGTTGCATC 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 703 GCTTTCGTGATTAAGTTCTTGCACACACTCGGCCAGTTGATGGGCTCTCGCAATC 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 480 TTAAGAACTGACATTGCTCATTTTATGACTTGTGTAACATATGGGAAGCGTATG 539
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 763 ATGAGGATTTGGAGTTTCTGTTGTTCTATGTCATGAGCTTTGGGCTAGTGTGTC 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 540 CTATCTTAATGTTGCGGATTTGCTATGAATTAACAAAATCCAGAACCAAGCGT 599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 823 GTTTAGTTCTTCTTGGGGATTTTGCACACGATTTACACAGTTGACAGCAAAAAG 882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 600 TTCTACGCTCTTTTGGGATTCGAGCTAATATTTCTTAAGCTTGTGCTGTCAT 659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 883 TTCTATCCCTGTTTGGGCAATGTTGCACTATCTCTCAGAGAAGACTGTG 942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 660 GTTTGGGCTTCAAGTTGAGACTTCCCTTCTGAAGTGTGATTCCTGGGGAATTTCT 719
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 943 AATATTTCTTAATGTAAGAAAGATCTTGTCTCGAGGATGAGCGGGCTGTTTCA 1002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 720 TTACGCTTTTGGGCTATGACTATTTGATCTGACTTGTCTATGCGCAGTTACTG 779
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1003 TTAAGAAGCATATGATGTTTGTGGGGATGGGCTCGCATCTGTTTCTCTACGTG 1062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 780 TCGATCATTAAGACGATTAAGCCGATCTCCCTCTATATTCAGAGAAGAAATGCAAAAG 839
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1063 TGGGGAATAA-----GATATGTGCCCTTCCAAACCCGTTAGC 1098
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 840 GGAAGAAAAGGTGTAACCTAAATGATATGAAGAAGATAGCTTCTCATCTGTATGA 899
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1099 AAGAAGAAAGGTGAACCAAGATGGAGACAAATGCAAGCTTGAAGTTCTTGCTGTA 1158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 900 TCTCCCTAATATCTTTTATTAATCACTCTGATGTTATGCCATGATGATTTGCAATTA 959
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1159 TCACCATCATATTAGGAGATCTTCTACTTGTGGTGTGATATGAAATAGTATCAACCTT 1218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 960 ATTCAGATGACTTGGAAAAGTCAAGTGAACATGCAATATCTTAATATGATGATAGT 1019
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1219 GTTGAAGTCACATGGAATAATCAAAAGCTTAAAGTCAGTTCCCTAGCCCAAGAAATATCA 1278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1020 GAGTTCATCGGGAAGACTTCTCTTGTGAGACTGGCCGATAGTATCCGTAATATCATGCTATT 1079
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1279 GCAATTAATGAGGACCTTCTCAACCTGACAGGATATTTGCAAC--ATTCACATGATGCT 1335
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1080 GTTGTGGAACGATATTCGTAATTTGGATAGCTTAACGTGAGCCCTAGTCACTCTGTC 1139
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1336 CTAAAGCAATACGTGTTTAAGAAGATATGTTGGGAGATAGCTGCAAAAGATCACACCAAC 1395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1140 ATGATTCCTTAACAGATATGCTTTCTGCTGCTGTTATCTTAAGAAAACCAAGCTCT 1199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1396 GTTCTGCTATTAACGGGTGCGCTTCTCTCTGATACGTGTTGGGGCCCATTTGCGA 1455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1200 GGGCTGCTGCTATGTTGGGTCAACCTCTCTCATGCTAGCTGTGTTGTGGAGCTATA 1259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1456 CCATTGGTTGCCAACCTTGGTATGACACCGCTACCTGCGACGATGCTAGCTTCCCTCA 1515
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1260 CAGAATATCTTTCGAATAATCCACAATAATACGCTCTCTTGTGACTC--AACTAAGAAATG 1316
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1516 GA-AGTATCTTACGAAGAGTCCAAAGTACAGCATTTCTGACTCCCTGCAATGCAAGA 1574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1317 GCCTATATCCCTTGTACCAAGAGCAAAAAGTCAAAAGGTAGAGCTGATTAATGATAGT 1376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1575 ATGCTATATTCATTGATGATGAGGTCTCCAAAGGTAAGCAAGCTGCAATGATGCTG 1633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1377 GCGCGCCGCTTGGAAAATCAGAGAGAGCTTAAATCCAAACAAGTTGCTCTATCTGT 1436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1634 -CTGAACCATTTGGGAATACAGCGGTGCTTAATCCAGCATTAATGATCTTACATTG 1692
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1437 GGAATATTTGAGATGACCCCTTATCTTGAAGATCTTCTTTCATCAATGCTATT 1496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1693 GGCTACATCGCCAAATTCACACACTTACCTTGAAGTCAATCTCTGTGGTATGACTGCA 1752
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1497 TGGTGGTTTTCGCACTAAGTTAAACAACAACTATTCTTA 1535
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1753 TCGTTAGCAGCAGCTAATGCTGAGGAGCACCAATTAA 1791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 26  
AF087957 1505 bp DNA Linear SYN 07-NOV-1998  
LOCUS Synthetic construct ATP/ADP translocase (tlc) gene, complete cds.  
DEFINITION AF087957  
ACCESSION AF087957  
VERSION AF087957.1 GI:3834578  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
artificial sequences.  
1 (bases 1 to 1505)  
REFERENCE  
AUTHORS  
TITLE  
Gene synthesis, bacterial expression and purification of Rickettsia prowazekii ATP/ADP translocase  
Prowazekil ATP/ADP translocase  
Unpublished  
2 (bases 1 to 1505)  
REFERENCE  
AUTHORS  
TITLE  
Membrane topology of the Rickettsia prowazekii ATP/ADP translocase  
revealed by novel dual pho-lac reporters  
Unpublished  
3 (bases 1 to 1505)  
REFERENCE  
AUTHORS  
TITLE  
Direct Submission  
Submitted (30-AUG-1998) Microbiology and Immunology, University of  
South Alabama, LMB Building, Mobile, AL 36688-0001, USA  
JOURNAL  
FEATURES  
source  
1..1505  
/organism="synthetic construct"

```

/db_xref="taxon:32630"
/note="coding region for Rickettsia prowazekii ATP/ADP
translocase engineered to optimize expression in E. coli
and to introduce unique restriction endonuclease sites in
every region coding for a potential hydrophilic loop of
(except loop 11) thus creating a cassette version of the
gene"
4..1500
/gene="tlc"
4..1500
/gene="tlc"
/function="translocase"
/note="obligate exchange system"
/codon_start=1
/transl_table=1
/product="ATP/ADP translocase"
/protein_id="AAC72100.1"
/db_xref="GI:3834579"
/translacion="MSTSKSENYLSELRKIIPLEYENKKFELPLAFMFCILNYSF
LRISIKDFVYVDIGTESISELKYIVLPSAVIAMIVKLCIDILKQENVEFYITSEFL
GYLAFVAFVLPYDPLVPHKITESLSLAYPNKFWIKIKYKWSFASFYTLAEIMGT
MMLSLFMQFANOITKIAEAKRFYMFGLANLALPVTSVIGVLEHKOIVAEHK
FVPLFVIMITSSFLIILTYRMNKNVLTDPRLPALVEKKTKAKISPIESKMITF
SKYGYIALILIAAGSVNLVEGYMKSKVELYPTKRAYITIMQOPFYQGVIAIHM
LIGSNILRKVSMLTAAMITPLMFTGAAFSEFEPDSVIAIMNLGLASSPLTAVM
IGMLQNLVSKGVKISLEDAIKNMAVYPLDRLRVKGAAGAEVIGRGLKSGGALIOST
FFILFVPGFEIAPYFAISIFIIIVILMIRAVKLNKEYOVLVKNKEK"

BASE COUNT      346 a      389 c      297 g      473 t
ORIGIN
Query Match      16.8% Score 274.2; DB 12; Length 1505;
Best Local Similarity 52.0%; Pred. No. 6.7e-55;
Matches 783; Conservative 0; Mismatches 688; Indels 36; Gaps 6;

66  GAAAAACCTTTTGGAAAAATTTGGCTCTTCTTGTCGGCATACATACACAGAGCTAAG 125
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
22  GAAAACTACCTGTCTGAACCTGCGTAATCATCTGCGCATGCAACAGTACGAAAAACAA 81
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
126 AAGCTTTCGCCAATGTTCTTAATGTTCTGATATTAACATTAACATTAACGTTTACGC 185
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
82  AAATTCCTCGCGTGGCTTCATGATGTTTCGATCCTGTCGAACTACTCTACCTGCGCT 141
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
186 GATACAAAAGACACTTATTTGTGGAGCTCTGCTGTTCTGGTGCAGAGCAATACCTTTC 245
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
142 TCTATCAAGAGGTTTGGTGT-----TACCGATATCGTACCGAATCATCTCTTC 195
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
246 ATCAAGTTTGGCTTTGTGTCCTGCTGCTAATTTCTTATGCTTATTTATGCAAGTA 305
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
196 CTGAAAACCTACATCGTTCTGCGCTGCTGCTATGATGATCATCTACGTTTAACTG 255
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
306 AGTAAATATTTTAAGTAAGACAGGCTTATTTATGACATGAGGAGCCCTTTTAAATTT 365
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
256 TGCACACCTGTAAGAAAGAAAGATATTTACGTTATCATCTTCTTCTCTGCGTTAC 315
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
366 TTTGCGCTTCCGACATGATTTATTCGCTACCGCATGTTTATCATCTACAGAAATTT 425
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
316 TTGCGCTGTTGCGTTCCTGTTCTGTCACCGCATCCGAGCTGTTACCCGACACAAA 375
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
426 GCTGACCGTTTACAGGCACTCTACCTCCAGATTTGTAAGACTTGGCCATCTTAAGA 485
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
376 ACTATCGAAAGCTTGTCTGTGCTTACCCGAACCTTCAAAATGTTTCAATAAATGTTGCT 435
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
486 AACTGACATTTGCTGCACTTTTATGTAATGTAATGTAATGTAATGTAATGTAATGTA 545
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
436 AAATGCTCTTTCGCTTCTTCTTACATGATGCTGTAACGTGAGGACCATGATGCTGCT 495
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
546 CTATGTTTTCGGGATTTGCTAATGAATTAACAAATTCACAGAGCAAGCGTTTCTAC 605
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
496 CTGCTGTTTCGGGATTTGCTGTAACAGATCAATAAATCGCTGAAGCTTAAACGTTTCTAC 555
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
606 GCTCTTTCGGTATCGGATTAATATTTCTTACAGCTTCTGTCGCGCAATGTTTGG 665
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
556 TCTATGTTTCGCTGCTGCGCAACCTGCTCTGCGGTTTACCTCTGTTGTTATGCGTTAC 615
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||

```

```

666 GCTTCAAGTTGACAGCTTCCGTTTCTGAAGGTGATGATCCCTGGGGAATTTCTTACGT 725
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
616 TTTCGACACGAAAAAAGCT-----CAGATCGTTGCTGAACACCTTAATGTTGCT 663
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
726 CTTTTCGATGCGTACATTAATGATCTGAGCTGTTCTTATGACCATGTTACTGATGATC 785
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
664 CCGCTGTTTCGTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 723
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
786 AATAGAACGATATGACCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 845
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
724 AACAAAGAGCTTCTGACGACCGCGCTGTGACGACCGCGCTGTGTTAAAGAAAAAAA 783
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
846 AAGGTCGTAACCTTAAAGTAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 905
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
784 A-----CTAAGAGTAAAGCTGTTTATGATGATGATGATGATGATGATGATGATGAT 837
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
906 TATATCTTCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 965
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
838 TACGTTGCTTACATGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 897
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
966 GTGACTTGGAAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1025
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
898 GGTGTTTGAATATCTAAAGTTAAAGAACCTGACCGACCAAGAGGCTTACACCATCTAC 957
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1026 ATGGGAACTTCTCTCTCTGACGCTGCTGATGATGATGATGATGATGATGATGATGAT 1085
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
958 ATGGGTGATGTTCCAGATGCTTACACAGGTTGGGTGCGTATGCTTTCATGCT---TGATGCGT 1014
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1086 GGTACGCTATTCCTTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1145
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1015 TCTAACATCTGCGCTAAAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1074
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1146 CTCCATACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1202
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1075 TTTCATGACGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1134
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1203 CTGCTGCTGATGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1262
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1135 CTGACCGGACATCTGCGCTTCTTCTGCGCTGACCGCTGCTGATGATGATGATGATGAT 1194
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1263 AATATCTTTCGAAATGACAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1322
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1195 AAGCTTCTGCTTAAAGGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1254
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1323 ATCCCTCTTTCGACCAAGCAAAAGTCAAAAGGTCAAAAGGTCAAAAGGTCAAAAGGT 1382
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1255 ATCCCGCTGACCAAGCAAGCTGCTGATTAAGGTCAAGCGCGGTTGAAGTTATGCGTGT 1314
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1383 CGCTTTCGGAATGACGAGAGGCTTTAATGCAACAGGTTGCTGTTATGCTGGAAGT 1442
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1315 CGTCTGAGTAAATCTGCGGCTGATATCATGATGATGATGATGATGATGATGATGATGAT 1374
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1443 ATTGAGAGCTAT-----GACCCCTTATCTTTCAGTATGATGATGATGATGATGATGAT 1486
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1375 TTGCGTTTCATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1434
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1497 TGTGTTGTTTTCGCAACTAAGTTAAACAAATATTTTAAAGGAGCTGCTTAAAGAA 1556
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1435 TGGATCTTCGCTGTTTAAAGCTGTAACCAAGATATCCAGGTTCTGTTTAAACAAAAGGAA 1494
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1557 CAGAGAG 1563
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
1495 AAATGAG 1501
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||

```

RESULT 27  
 AY120885  
 LOCUS  
 DEFINITION  
 (nnt) gene, complete cds.  
 ACCESSION  
 VERSION  
 AY120885.1 GI:22035402





RESULT 28  
AE001323/C  
LOCUS  
DEFINITION Chlamydia trachomatis section 50 of 87 of the complete genome.  
ACCESSION  
VERSION AE001323 AE001273  
KEYWORDS  
SOURCE Chlamydia trachomatis.  
ORGANISM Chlamydia trachomatis; Chlamydiaceae: Chlamydia.  
REFERENCE  
AUTHORS 1 (bases 1 to 24297)  
Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,  
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
Koonin, E.V. and Davis, R.W.  
Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis  
JOURNAL Science 282 (5389), 754-759 (1998)  
MEDLINE 99000809  
PUBMED 9784136  
REFERENCE  
AUTHORS 2 (bases 1 to 24297)  
Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,  
Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.  
Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE  
AUTHORS 3 (bases 1 to 24297)  
Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,  
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
Koonin, E.V. and Davis, R.W.  
Direct Submission  
Submitted (20-May-1998) Program in Infectious Diseases, University  
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
FEATURES  
Source  
Location/Qualifiers  
1..24297  
/organism="Chlamydia trachomatis"  
/strain="D/UW-3/CX"  
/db\_xref="taxon:813"  
complement(83..1705)  
/gene="CT495"  
complement(83..1705)  
/gene="CT495"  
/codon\_start=1  
/transl\_table=1  
/product="ADP/ATP Translocase"  
/protein\_id="AAC68096.1"  
/db\_xref="GI:3328932"  
/translation="MSSEVKSFSKRGYFPPIYVSERSKFIPLFLAFVGVNVALK  
TTKDSLVLVSGRAGAETVPLKVGIVGAVIVMIGWMSRRYSRGIVFTSLVGGL  
GFALFTVIVIPIDALHLNKLAKLOSILPPGGRGVNVOVWSYSLIYVMSLMS  
IVSLTFMGVANHITSVREAGRFYALINTGNTSSVFEVSLMGRNPVLAFFMAVD  
PMHEVLNITLLIYLAGGVILYLOKDLMDETSMLKEGLAEMSVANOLKKEKRSK  
AKASIFALLIRSYILGIAVIVSYNVIHLFEVVKDQCRITYASVENSMSRT  
TTLTGIVSALTGIFAAGOTIRKMGWTTGALPLPTMLITGALFEGALYAVAGDMITG  
GILGISPLVITLAVLGVOVNEFSRAIKFYFDQTEMAFIPLEDEKNKGAIDGVIS  
RVKSGSLVYOGLLIFSSVAASIMNTIVLLALGSMIFIVIAMLGREVYAKTEALF  
RVNSEDVLOEEREASIVDAESREBEVTVL"  
complement(1917..2423)  
/gene="pgsA\_1"  
/note="CT496"  
complement(1917..2423)  
/gene="pgsA\_1"  
/codon\_start=1  
/transl\_table=1  
/product="CDP-diacylglycerol--glycerol-3-phosphate  
3-phosphatidyltransferase"  
/protein\_id="AAC68097.1"  
/db\_xref="GI:3328933"  
/translation="MROFCNLLSLRWALLFCOERTITRLVITFAAMVSDVLGYL  
ARRYNATSRIGSITLDPADNKTIFLLIGVGLFWENSLGITHALIFSRDILPVEFGFL  
SWVGMKGYDYRALSFQKFTVVOFLLFGNTIGMEKAPVLMLAPLVILGALYFLERV

1  
gene  
complement(2610..2693)  
/gene="trnA-Ser-2"  
complement(2610..2693)  
trnA  
/gene="trnA-Ser-2"  
/product="trnA-Ser"  
complement(2819..2968)  
gene  
/gene="CT496.1"  
complement(2819..2968)  
CDS  
/gene="CT496.1"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAC68130.1"  
/db\_xref="GI:3522901"  
/translation="MGNGFFVVAIVRCLFLMLVKALASRLGIRGITPOBERGEEK  
YRPLG"  
2937..4355  
gene  
/gene="gnaB"  
/note="CT497"  
2937..4355  
CDS  
/gene="gnaB"  
/codon\_start=1  
/transl\_table=1  
/product="Replicative DNA Helicase"  
/protein\_id="AAC68098.1"  
/db\_xref="GI:3328934"  
/translation="MANTPKKPPDPTLLSPNSKEEMIVICMLTSVNHNLNANIL  
QEDDPFLEHRIIRVLODARKSDRPNDPHLTGEBLKRDDLTNTGGPSYIITSEFA  
GTSVITEYATIRKSTLRKMIQAKTEKKAEEPRDVTALDQONLFRITSSOT  
MLAPHLVADKLKVAASSKDKSFLIALQERKAEQASAHDSSEMLSGFPPLDLDRK  
NISGEFSPNLITILARPMAGKTALANTVNEFCDSRLPVGIFSELEMTVDLIRIIC  
SRSEVAKKISVDISGRDFORVSVVMEHEFTLLDDYGLKITDLARARAKMS  
YDIOFLVIDYQILSSSGLNLSNDSRNOEISEIRMLKNLAREINIPILCSLSRY  
EDRANHRRPMSDLRESGSIEDQADQIMFLRLREYIDPRDKGTALILVAKNRHSIS  
VOLYERKDFARFRNACGEFPG"  
4649..6481  
gene  
/gene="gIda"  
/note="CT498"  
4649..6481  
CDS  
/gene="gIda"  
/codon\_start=1  
/transl\_table=1  
/product="FAD-dependent oxidoreductase"  
/protein\_id="AAC68099.1"  
/db\_xref="GI:3328935"  
/translation="MTTFPPVDYDVIIVIGAGGCEAAVCAAKMGASVILLTNSLDIT  
KLSCNPAGVGIGKGIYREIDALGSIAGETDLSGIOFRIITNOKGAVRAPROVDR  
OLVITHMKRLLEQVPGHITWGTAEALDNGEYLVGVSTKEGAVLGTATVVLSSGTFM  
RGLIHITQNFSGRGLDMAASLGISEDLRKIDGFLKLTGTTPARLASSIDFVME  
QPGHNACFVRNEMFVPTLPVSGCHITRTDQPLKLTMLHRSALVGGRIEYGVGR  
AIPSIRIEKIVFAADKRNHIFIEPEGNTQCVYVNGISTMPPVOVDITRSVGLNA  
AIIIRPVAIYEDVHGVNIPPSLESKLIEGLFLCGOINGTGYEEAAAOGLIGVNA  
HVKSLIRHPPEVROESYIGWMLDITTOVDEPFRMFTSAEHRLLTRDNACMRIS  
HYGSLGLISSEIRAMQEOKACIEQEKERSKTFKRYGDIYVPLTYLCEPEVSYO  
LTFEPADVRDLPVGVASLEMETIKSYISRQDTLIRMSRSENISIPEDIDVHSIS  
ALSLAEKSKFTPTRTIGSAARISGISVADIOVLWVSLKKDAH"  
6471..7172  
gene  
/gene="lplA\_2"  
/note="CT499"  
6471..7172  
CDS  
/gene="lplA\_2"  
/codon\_start=1  
/transl\_table=1  
/product="Lipote Protein Ligase"  
/protein\_id="AAC68100.1"  
/db\_xref="GI:3328936"  
/translation="MLNCFVHCESGPIRKQQLQEBALRTSSQNCFLVNLTPLEAV  
VLGISRKPERDLHVEHLKEDGIPITIRRYSGGTVFLDADSLMWSMIINSPTPSSKD  
LLOWTODIYAPITPTGKTIENDVITPDKRTKGAAOYIOKRYRWVHTTFLMNNPKKL  
ARYLPPPEIOPSVYKSHDEFLTITIELDPSRDFLSQLKQSAASKMWEGDSIGTL  
TPMISLPHRRATQIL"  
complement(7229..7654)  
gene



```

CDS
    /gene="ndk"
    /note="CT500"
    complement(7229..7654)
    /gene="ndk"
    /codon_start=1
    /transl_table=11
    /product="Nucleoside-2-P Kinase"
    /protein_id="AAC68102.1"
    /db_xref="GI:3328937"
    /translation="MEORTLSTIKPDSVSKAHGELIALFEPSGLRIAMKRVHLSYKE
    AEGRYVHKERPEFOELVDENISGFVVMVLOGEAVVRNRELKMGATNPKAEBSIR
    ALFESIGVNAVHSDSLNNAIEVSYPEFAETIVNSVA"
    complement(7814..8416)
    /gene="ruva"
    /note="CT501"
    complement(7814..8416)
    /gene="ruva"
    /codon_start=1
    /transl_table=11
    /product="Holliday Junction Helicase"
    /protein_id="AAC68102.1"
    /db_xref="GI:3328938"
    /translation="MYEIKGTLTHINESYVIESFGIYATIMSERPLVDLRPMHQ
    EVLIYVSHVIRETEHVLXGFSRARECEFRLLISGIGPKGLSLNMFPLQELCSI
    ARLENVKAIASVPGIGKTKAEKLMVDLQKLPFLLEEVVPSSTANSEFKEGIG
    ALMNLGFRILADRMTEAVKELSEASVAVELLPLAKKS"
    complement(8436..8948)
    /gene="ruvc"
    /note="CT502"
    complement(8436..8948)
    /gene="ruvc"
    /codon_start=1
    /transl_table=11
    /product="Crossover Junction Endonuclease"
    /protein_id="AAC68103.1"
    /db_xref="GI:3328939"
    /translation="MADLMGIDPSTLVGYALIKENRYHIIHPHSGKVKLSOKLAT
    AHRYKQLEIESTILOESPRAVVLQYVILQHPQSTIKLGARGLVLLASLOQV
    FEYAPNTAKKAAGVGNMKSQVOIAMSKLRLVPLLAEEDNEDIDATRALAMCHHLA
    PYQDKRTLV"
    complement(9056..9610)
    /gene="CT503"
    complement(9056..9610)
    /gene="CT503"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAC68104.1"
    /db_xref="GI:3328940"
    /translation="MWMSLVKRLRLSLRFRKVRAPTSPODYEYSRILOFLMK
    DSDIYEMTEOMLEFGASKMSEDLLRKKKMOIOSAKEOSILAMBOOVCEIKSOLA
    SOENANOEISKLQAEHNHLORELAEKIQARHNDVIDELKRDVLESVOQMEVSEGR
    RUCLEHKIRLVEQIDRFLAKEV"
    complement(9695..9777)
    /gene="tRNAleu_3"
    complement(9695..9777)
    /gene="tRNAleu_3"
    /product="tRNA-Leu"

Query Match      15.7%  Score 256.4:  DB 1:  Length 24297;
Best Local Similarity 49.7%:  Pred. No. 9,3e-51;
Matches 749;  Conservative 0;  Mismatches 736;  Indels 21;  Gaps 3;

```

```

Db 1578 TAAGACCAAGAAAGACCTCCCTGTTTATGTCGGCTGAGAGCAGGGGGAAGTCATAC 1519
QY 241 CTTTCATCAAGTTTGGCTTTGTCGCCCTGTCGCTATTATCTTATGCTTATTTATGCA 300
Db 1518 CTTTCATCAAGCTGGGGGATGTCCTGGGGCCCTTATCGCACATGATGTATGAGAT 1459
QY 301 AGCTATATATATTTTAAAGAACAGCCCTTATTTATGACAGTGGGAAGCCCTTTTAA 360
Db 1458 GATGAGCTGAGCTTATTCAAGAGTACCCTGTTTATCTTATGTCGGCGGTTTTTTAG 1399
QY 361 TTTTCCTTGGCCCTGTTCCGCGCTGTAATTTATCCGCTACGCGATGTTTACATCCACG 420
Db 1398 GCTTTTTCGCTGTTCCGCGACAGTATTTATCCATAGAGATGGCGTCACTAATAA 1339
QY 421 AATTGCTGACCGCTTACAGGCCATCTCTACCTCCAGAGATTGCTAGGACTCGTTCCATCT 480
Db 1338 AATTGGCGGCAAAACATCAAGTCATCTCTACCTCGGGGGGAAGAGGTTTGTATGATG 1279
QY 481 TAAGAACTGAGACATTTGCTGCAATTTATGACTTCTGTAACATATGGGGAAGGTCATGC 540
Db 1278 TTCAATACCTGAGACTATAGCCTGATATGTATGTATGTCGATGATTTAGAGTTCCATAGTTT 1219
QY 541 TATCTCTAATGTTCTGSGGATTTGCTAATGAATTAACAAAATCAAGCAAGCAAGCCGTT 600
Db 1218 TGTCTACTGCTGTTTGGGGGAGTCCACACCATATTACGAGTGTTGCGAAGCAGGGCGGT 1159
QY 601 TCTACGCTCTTTTCGATTCGAGCTAATATTTCTTACTAGCTTTCGTCGCAATTG 660
Db 1158 TTACACCTCTTATTAATACGTGATTAACCTCTCTCTGTTTTCGCGAGAGGTTTCTT 1099
QY 661 TTTGG---GCTTCAAGTTGAGAGCTTCGCTTCTGAAGSTGAGATNCGTTGG----- 710
Db 1098 TGTGCTCGGTAGAAATCTCTGATATTCCTTCCCTATGCGCGTAGATCCCTGCGATGAG 1039
QY 711 -----GGAATTTCTTACGCTCTTTGATGAGCTATGACTATTTGATCTGCACTTGTTC 762
Db 1038 TGTGCTCATATACACTGACTATGCTGCTGTCGCGCGAGTAATCTCTATCTATATAC 979
QY 763 TTATGGCCAGTTACTGCTGATCAATAGAACGTAATGACCATCTCCCTCTATTAATC 822
Db 978 AGAAGTTAGATCGCTTGATGATGATGAGACGTCATCTTAAAGAGGGGTTGGCAGCAAA 919
QY 823 CAGAAAGAAATGCAAAAGGGGAAAAAGGTCTAAACCTTAAATGATATGAAGATAGAGCT 882
Db 918 TGTGCGTCCGCTCAGCTTAAGAAAGCAAAAAGCGGTGCAAGCAAAAAGCAAAAGCTCT 859
QY 883 TCCCTATCTTGTATGATCTCTTATATCTTTTATTAACCTCTTGTGTTATGTCCTATG 942
Db 858 TCGCTCTCTCCCTCCGCTCCGCTATCTATTAGCATTGCGCGTCTGATCTCTTACA 799
QY 943 GTATTGCAATTAATTAATCGAAGTACTTGGAAAAGTCAAGCTGAATGCAATATCTTA 1002
Db 798 AATTACTCATTCACCTTTTGAAGTGGTGAACCAACCAAGTCCTCGGATTTATGCTT 739
QY 1003 AATGATGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1062
Db 738 CTGAGATGATATTTATTTCTTATATGATGATGATGATGATGATGATGATGATGATG 679
QY 1063 TACTTATCATGCTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1122
Db 678 CTTTGACAGGATTTTTCGCTGCTGGACAAACATTCGCGGTTGGGATGAGCATTTGGGG 619
QY 1123 CCTTACTACTCTGTCATAGTTCTCTCTAACAGATATCGTTTCTT---CGCTTGTGTA 1179
Db 618 CCTTGGTACCCCTTAAACATGTTGATTAACAGAGACCTTGTCTTGGCGGCTATATG 559
QY 1180 TCTTTGAACCAACAGCTTTCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1239
Db 558 CTGTAAAGGAGATGCTATGATTTTGGCGGATCTTGAAGATCTCTCGGATGATCTTA 499
QY 1240 CTGTGTTTGGAGCTATACAGAAATATCTTTCGAAATTCACAAATATAGCTCTCTTGG 1299
Db 498 CTGCTTGTAGAGGTGTTCAAGAACGTCCTTCGAGACCAATTAAGTTTACCTATTTGG 439

```

QY	1300	ACTAACACTAAGAAATGGCCCTATATCCCTCTTGACCACAGACGAAACAAAGTAAGG	1359
Db	438	ATCCAAACCAAGAGATGGCATTTATTTCTTTGGAAAGATGAGAAAGATTACGGTAAGG	379
QY	1360	CTGCTATTGATGTACATTGGCCGCCCTTCGAGAAATACAGAGAGACCTTTAATCCAAAG	1419
Db	378	CGGCTATTGATGGGGGATCTCCAGAGTAGGGAAGCAGGGGGCTTTAGTTTACCAAG	319
QY	1420	GTTTGCTGCTTATCTGTGGAGTATTGGACGTATGACCCCTTATCTTGCAGTGATTTTC	1479
Db	318	GACTCTGATCATCTTCTGCTGCTGGACGTAGCTTAATATCGAATTACGATTGTGTGT	259
QY	1480	TTTTCATCATCTGCTTATTTGGTGTGGTTTCGCAACGTAAACAAAGATTTCTTAGCC	1539
Db	238	TGCTGCTTTTAAAGGAGTTGGATCTTTGTGATCTGCTTGGTTAGTAGAGATACACAGCCA	199
QY	1540	AGTCTG 1545	
Db	198	AGACCG 193	
RESULT 29			
LOCUS	TAJ10587	1623 bp	DNA linear BCT 23-MAR-1999
DEFINITION	Chlamydia trachomatis npt2 gene.		
ACCESSION	AJ010587		
VERSION	AJ010587.1	GI:4538861	
KEYWORDS	npt2 gene, nucleoside triphosphate transport protein 2.		
SOURCE	Chlamydia trachomatis.		
ORGANISM	Chlamydia trachomatis.		
REFERENCE	1. (bases 1 to 1623)		
AUTHORS	Tjaden, J., Winkler, H.H., Schwoppe, C., Van Der Laan, M., Mohlmann, T. and Neuhaus, H.E.		
TITLE	Two nucleoside triphosphate proteins in Chlamydia trachomatis, one for net nucleoside triphosphate uptake and the other for transport of energy		
JOURNAL	J. Bacteriol.	181 (4), 1196-1202 (1999)	
MELINE	99138740		
REFERENCE	9973346		
AUTHORS	2 (bases 1 to 1623)		
TITLE	Neuhaus, E.		
JOURNAL	Submitted (25-AUG-1998)		
FEATURES	Submitted (25-AUG-1998) Neuhaus E., University of Osnabrueck, FRG Biology / Plant Physiology, Barbarastr. 11, D-49069 Osnabrueck, FRG		
source	Location/Qualifiers		
gene	1. .1623		
	/organism="Chlamydia trachomatis"		
	/db_xref="taxon:813"		
	1. .1623		
	/gene="npt2"		
	1. .1623		
	/gene="npt2"		
	/function="nucleoside triphosphate uptake"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="nucleoside triphosphate transport protein 2"		
	/protein_id="CA839535.1"		
	/db_xref="GI:4538862"		
	/translation="MSSEVKSFSKFRGTYEPPIYRSEPSKFIPLFLAFYGVNVALLK		
	TTKTSILVGSRAAGAAEYIPELKVAGIYGAIVYIMYIGMSRRSRGCTVETLSYGGTL		
	GFPLPRTVYIPIGIDALHLNKLAKLADSIIPGGRGVVNYQYISLYIYMSLMS		
	VLSITLFWVANHITSYREGRFPALINTGLNLSVPAGEVSLMLGSPVIAPFMAVD		
	PWHEMLNLTLLIYLAGVILYLQKLDRLMDERSMLEGLAEMVAOLKKEKRSK		
	AKAKSLPALLRSRYLLGIADVVLVYLVIHEVWKDDVCRIYASVEFNSYMSRI		
	TTPLGIYALAGIAPAGOTIRBMGWTGVALVPLITIIITGALFGALYAVKGDAMIRG		
	GLIGISPLVLTAMGIVQNYEFSRAIKETFTDQTEMAFPLEDEKRYKGAIDGVIS		
	RVGSGGSLYOGIILITFSSVAAASLANITIVLLALAGLSMIFIVIAMLREYTAKTETLV		
	RVNASEEDVYLQEEELASVLDASERREPVTLL"		
BASE COUNT	374 a	313 c	410 g 526 t
ORIGIN			

[illegible]

gene  
CDS  
/gene="RP226"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA14689.1"  
/db\_xref="GI:3860789"  
/translation="MLNKLCDLFLNLINLLVTVGVASPPPLPSIP1IADVTPDKMVISNSISFEKPFQKOKKRNISQHEQOTAIHOESQIDISRELNEQSEPTIDGSTITLSVASVNSIDSKAEYENSTNLAASVNTDIDVQKQDEPVSASEP1DIGNKPTLSANHNEEAVASNDKEENLTNS1ITPVVPSVPVIS1PRAQDVNVVSOOSVOJYKPTNLTS1RNPI1LHHTHDTLANKVEKNLESTISINPTTNSVPS1DQ1QTLTNLVPATLTHVSVSTVYHNSHNSAOPIRT1SNFVETSSVYLRATESVP1NNSQELFVSESTKQDMWYPI1MVL1VVDPNK1SQR1LAE1QKNNQD1INNOAESHSVSSVY1QKQNDKVNANRTSESTKEFYKQETQML1FLPDD1V1GK1LEAT1QDMHGY1KFTQKEEWM1NAER1KLVSEF1KYDD1DNKNKQ1YAN1LS1S1VADNA1FAVER1NN1LE1RLAL1DVA1PLQAK1NSTG1E1LTS1YCN1Y1LAK1ELV1RG1TS1VLK1NDEK1Y1PL1DAL1AKGN1Y1CML1MK1AK1XN1"  
2025. 4013  
/gene="RP227"  
2025. 4013  
/gene="RP227"  
/codon\_start=1  
/transl\_table=11  
/product="DNA GYRASE SUBUNIT B (gyrB)"  
/protein\_id="CA14690.1"  
/db\_xref="GI:3860790"  
/translation="MSDLEFSLNKKRNKYVTYNSAKDIEVLQGLEPYRKRNPMYIGGSDSNMHLVSEVLNDAMDEAVAGFAS1IMIKMODHS1TTFDMGRG1PIDNHRKPGSKALV1LTHHS1GSKPNNYVH1SG1LHG1GV1SVANA1SKHFK1KYKQK1LV1SOSY1SKG1AL1T1LH1S1EAK1SK1RLRG1TS1NPT1DPL1E1SK1LH1NPK1Y1E1AR1SK1Y1LRG1S1IMBECEVE1PSD1TRK1KAL1N1F1NGL1D1YSK1S1LDN1V1P1E1T1SG1NES1VD1K1LE1WA1ICQND1TSA1E1MOS1YCN1V1P1DQ1G1H1EG1K1S1AL1R1G1K1A1S1EM1GN1K1R1AN1LT1T1ED1LE1T1A1E1R1S1IF1E1P1E1PS1FQ1G1T1REK1LV1SG1V1SK1L1E1V1N1IK1RD1H1D1L1SS1DK1V1LA1TH1LE1HT1A1E1A1E1RR1K1N1R1N1R1S1K1SV1Q1K1L1P1G1K1L1AD1CT1P1SA1E1G1EL1F1E1B1E1DS1G1A1G1G1SA1K1R1N1D1E1O1V1A1PL1MG1K1V1LA1V1A1S1T1Q1L1K1R1L1INNO1A1D1E1I1A1G1S1K1N1Y1K1EM1L1R1E1K1I1M1TD1V1G1A1N1A1S1L1M1TP1E1P1L1RM1K1L1V1E1G1H1Y1L1AK1P1L1R1L1Q1N1Y1Y1A1R1CDEER1K1L1Y1R1K1LS1A1SK1A1K1E1V1G1R1F1G1G1E1M1P1R1Q1L1E1T1M1P1E1R1K1SL1K1V1TL1ED1VQ1NDK1IV1D1LM1K1K1R1P1R1Q1FY1EQ1ALV1K1MD1I1NKL1D1"  
4122. 5447  
/gene="RP228"  
4122. 5447  
/gene="RP228"  
/codon\_start=1  
/transl\_table=11  
/product="PAII-SPECIFIC PROTEASE PRECURSOR (ctp)"  
/protein\_id="CA14691.1"  
/db\_xref="GI:3860791"  
/translation="MLRFLTALPLSLNCAIEGKETEKISNQRYAKOFQDVERIEKDVYQVDDKQIMDEAINGLMSLDPHSNVYDEDELEDTFTTKGEGFQIGVEMYDSGGAAK1I1S1D1D1L1P1AF1R1AG1L1K1G1D1Y1V1G1V1D1E1L1V1T1R1K1E1K1P1G1R1V1L1I1K1I1E1E1A1R1P1O1E1L1E1R1E1V1K1I1K1R1A1H1E1K1N1N1Y1R1T1P1E1N1S1E1L1K1A1V1K1L1E1E1SK1DN1K1I1T1D1L1R1N1N1G1I1D1A1V1S1D1Y1T1G1S1V1T1Y1R1G1T1S1S1N1E1T1K1A1E1P1L1E1K1R1M1V1L1N1G1NS1A1S1E1T1V1A1C1A1D1H1K1A1I1L1G1R1K1E1G1S1V1O1L1T1N1S1E1A1A1K1L1T1SK1Y1T1P1S1G1R1S1D1A1C1E1P1D1L1E1P1A1K1V1P1E1Y1K1I1D1R1F1S1S1L1K1N1L1K1D1N1A1K1N1K1S1N1E1Y1T1K1N1K1Q1E1S1F1E1L1Y1K1Y1O1F1A1R1A1Y1V1T1G1L1I1N1T1L1E1Q1A1K1"  
5653. 7146  
/gene="RP229"  
5653. 7146  
/gene="RP229"  
/codon\_start=1  
/transl\_table=11  
/product="HISTIDINE KINASE SENSOR PROTEIN (bara)"  
/protein\_id="CA14692.1"  
/db\_xref="GI:3860792"  
/translation="MNNKRYLVRSLFILLPLNTIYINMLFYRYPMIKEMIIKOVALEHTK1V1E1L1Y1D1N1I1N1H1K1N1V1SK1L1K1E1D1Y1K1L1L1O1D1Q1I1N1V1K1T1A1Q1M1P1N1L1N1S1I1S1Y1D1L1K1N1K1T1S1N1M1L1H1M1V1N1Y1K1D1S1L1E1I1V1T1R1I1D1I1F1K1S1F1S1A1P1L1R1D1A1E1G1I1S1H1L1L1P1K1V1E1N1E1S1D1L1E1H1A1S1F1V1S1Y1P1V1D1N1D1P1D1A1V1E1I1N1T1S1Q1W1K1I1S1L1E1Q1K1V1T1E1I1F1E1I1T1C1T1I1S1N1Y1A1R1O1I1E1O1L1E1T1R1N1K1A1O1L1E1K1E1K1S1S1N1T1K1F1F1A1N1S1H1E1L1R1P1N1A1I1G1E1S1E1I1S1M1S1R1D1P1E1K1S1N1Y1K1D1H1A1G1K1H1L1S1M1I1N1D1L1S1K1A1S1A1DK1L1K1V1D1N1D1D1N1K1L1S1S1L1L1K1P1R1A1Y1O1E1V1L1S1R1E1H1Y1I1N1A1D1R1K1Q1V1L1N1L1S1A1V1E1F1G1S1G1V1T1S1L1E1K1E1D1A1K1V1Y1K1V1I1D1G1E1I1G1E1E1K1D1P1T1F1L1A1F1Q1I1D1E1L1S1R1K1E1G1E1G1L1P1L1R1K1V1E1L1M1G1K1F1D1Q1S1K1I1N1G1T1Y1V1T1P1K1Y1D1S1I1E1T1"

gene	7374 . .8126	/gene-"RP230"
CDS	7374 . .8126	/gene-"RP230"
		/codon_start=1
		/transl_table=11
		/product="unknown"
		/protein_id="CA14693.1"
		/db_xref="GI:3860793"
		/translation="MKLITLLTFLPSLFLSGSEMPQKPLKYAVNNDFENRLDEOF
		OETRLKIGVEYLLOHRIIDLKONLMLNOEENIEVLETFDEPKODIFDIALLEGHNDH
		QKCFPEVKKDIDAPYKQAYDILAAVYKONKLTFAKDRKFNFOKYPNNLSLSMAYFWW
		ACCFEOKQYVNAATINYLYCKYKESPKGAKSSDGLLKLALSLGELKKMOEAQNIILAKLE
		KEPPIKRTISYKMKTEDAKIKFGCKIKNNKLI"
gene	8123 . .8785	/gene-"RP231"
CDS	8123 . .8785	/gene-"RP231"
		/codon_start=1
		/transl_table=11
		/product="unknown"
		/protein_id="CA14694.1"
		/db_xref="GI:3860794"
		/translation="MTDLDEVSDONEEKRLLPEPKRLVPIIIITIVYTIWITIKN
		MKDNRKNNQKNGDILVKTIGLDTTRKDKAKALFNLEMLTSSNTRKKEITALEQVADH
		RMSKEVYCAKNNLNNKIIDNEKYSIETSTAYATACCIIVDDQSLNTQDKKLOKRYLNTA
		YVDDEKCPWATATYIIKALIDIKHNMKTQAEKNLKNLASNVSPLDKQAKALLVSL
		SK"
gene	8812 . .10059	/gene-"RP232"
CDS	8812 . .10059	/gene-"RP232"
		/gene-"RP232"
		/codon_start=1
		/transl_table=11
		/product="unknown"
		/protein_id="CA14695.1"
		/db_xref="GI:3860795"
		/translation="MKKKLALLLPILISGNGLSKSVKNIVDLTRPLVIOQNEPIY
		LDNSRKIPFVNNMLKNKOYSPAKSKMAIEPVIGDMYITADIRANISAFSEIKNKITII
		WYNNLSHKHKNDYIGCGILHHNGKLYIYGARLLVLDPAKSGYEIRRELDPIRIKRP
		IALNLTHTIYVOTISNOTIADSETLKTITMDHESIAELISTYSMIPIVOHNVIVYVNA
		TGOVYALNINKEGVKWNFEFTLNDHTAIPNPDSSILCTYHDSMNLVYATGGKILIT
		KINLSTGYVLMQINADIOSMSLIGNSLPIINMROIALNPETGKKEVADLVNERN
		DKRLKSTYFLPEFVGVDNNNORSLSNVISDGLYNFNIDSGLMKMPHIIKIIINIRY
		YGLRSNNLTPESTDQVIFGSGI"
gene	10279 . .10746	/gene-"RP233"
CDS	10279 . .10746	/gene-"RP233"
		/codon_start=1
		/transl_table=11
		/product="50S RIBOSOMAL PROTEIN L13 (rplM)"
		/protein_id="CA14696.1"
		/db_xref="GI:3860796"
		/translation="WKTYSAKPSELIEKKMWVIDAKNVYVGLRASVAIMLRGHRKPSFT
		TPHLDGCGNIITILINAEHILKLGKILNPKDGKVYVRHNGFPGGIKDTAGKILSGKYPEE
		RIYKAAYVRMIRTRNLGAKOMSNLYVYANCHEPHMAQOPIIYDPASENPKKK"
gene	10753 . .11238	/gene-"RP234"
CDS	10753 . .11238	/gene-"RP234"
		/codon_start=1
		/transl_table=11
		/product="30S RIBOSOMAL PROTEIN S9 (rpsI)"
		/protein_id="CA14697.1"
		/db_xref="GI:3860797"
		/translation="MTELIKTEKVVYKQLTRESKLSYLKIPREKIDSVSRYATGKRR
		NAIAVWMLKVGSGKIVYNNKILNIOYFEPSEYVVKIILPFILTKTIDQYDVLCYVKGSG
		ISGCGAILHGISKALDKSAPCFHAIKKGGLLTRDSVVERKKYGGQRKARKKQGFQSK
		R"
gene	11279 . .11352	/gene-"rRNA Met (CAT)"
rRNA	11279 . .11352	/gene-"rRNA Met (CAT)"

[illegible]

QY	730	TGATGGCATATGACATCTTTGATCTATGACATGGTCTTATGAGCCAGTAACTGAGTCAGTACATA	789
Db 309771	TCATGCTTCACGCAATCTATTCATCATTTATTTGTTACTGCGAGAAATTAATTCCTATGTTTTAT	309830	
QY	790	AGAACGATTTGACCGATCCCTCGCTTCTATTAATCCAGAAAGAAATGCAAAAGGGGAAAAAG	849
Db 309831	TTAGAAATATAAATAATTTATTTTAACTAATTTCTATTAATGTTTACATGTAAATAAG	309890	
QY	850	GTGCT-----AAACCTAAATGAATATGAAGAATAGCTTCTCTATCTTGATAG	900
Db 309891	TTGGCGCTAAAAACAACAAACAAACCTGCAATTAATGAAAGATATAAATAATTAATTCATT	309950	
QY	901	CTCCCTATATTCCTTTTATTAACCTCTCGTGGTATGGCTATAGTATTTGATCTTAACCTAA	960
Db 309951	CAAAATATATATAGCTGCTATTTGATTTATTAATTAATTCCTTTATGATTTACTAATTAATATAG	310010	
QY	961	TCGAAGTACTTTGAAAAAGTACAGCTGAAGAACTGCATATCTCAATATGATAGTACTAGTATAG	1020
Db 310011	TTGAAGAGACCTTTGAAACCGAAATAAAGAAATTCATCCAAATATCTAGATTTATGTTA	310070	
QY	1021	AGTTGATGGGGAACCTTCTCTCTTGAGACTGGCGTAGATATCCGACTATATCTATTTTG	1080
Db 310071	ATTATATGGGCAATGTTTATATTTTGAGATGGGATCTCATGTGTTACTTTCATGATATAA	310129	
QY	1081	TTTGGTGTAACGCTATCTGTAATTTGATAGTGTATACGAGCCCTAGTACTCTGTCA	1140
Db 310130	---GGTAGTAATAATCTTGAAAGGCTTGTTGGCTCATTTCTGCATTAATTAACCTCTAATTA	310187	
QY	1141	TGTTCTCTCTTAACAGATATCTGTTTCTTCGCGCTCTGTATCTTTTGAAGAACCAAGCTCTG	1200
Db 310188	TGTTATCTATTAACAGGCTTCACTGTTTTTATTTATTTTATTAATTTTGAAGAAATAGTA	310247	
QY	1201	GGCTGGTCGCTATGTTTCGGTACAACTCTCTCATGCTAGTCTGTGTTGTCGAGCTATAC	1260
Db 310248	CATGTTTGGTGATTT-----TATCTTCTATATGATGAGCGATATTTGTCGAGCAATTC	310301	
QY	1261	AGAAATATCTTTCCGAATATCCACAAATATACGCTCTCTTTGACTCAACTAAAGAAATGGCCT	1320
Db 310302	AGAAATATCTTTAGTAATAACGCTAAATATTTCAATTTTGATTCACAAACAAAGAAATGGCAT	310361	
QY	1331	ATATCCCTCTTGACCAAGACAAAGAAAGTAAAGGCTGCTATGATAGTTAGTTGGCG	1380
Db 310362	ATATTCCTTTATCTTTAAGACTGAGAACTTAAGGAAAAAGCCGCTGTAAAGTAATAGAA	310421	
QY	1381	CCCGCTTCGGAATATCAGAGAGAGCTTTAATCCAAACAAGTTTGCCTGTTATCTGT----	1436
Db 310422	CGAAATTTGTAATAACACTTGAGAGCATTTATCCAGCTCTTGATATTTATTTATTCGA	310481	
QY	1437	--GGAATATTTGGAGCTATGACCCCTTATCTTGACAGTATCTCTTTTCATCATTTGCTA	1494
Db 310482	CGCGTACCTTTGATCTTTATTAATATTTTACTGTATTTTAAATGATGATGATGATTT	310541	
QY	1495	TTTGGTTGGTTTCTGCACACTAAGTTAAACAAACATTTCTTGACCGAGCT	1514
Db 310542	TATGATTTTGGATATTTATTAATAATTAATTAAGGATATATTAAGCTGTGT	310591	
RESULT 31			
AE002346/c			
LOCUS	AE002346	12084 bp	DNA linear BCT 26-Mar-2000
DEFINITION	Chlamydia muridarum, section 74 of 85 of the complete genome.		
ACCESSION	AE002346 AE002160		
VERSION	AE002346.2 GI:8163321		
KEYWORDS			
SOURCE	Chlamydia muridarum.		
ORGANISM	Chlamydia muridarum		
REFERENCE	Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.		
AUTHORS	1 (bases 1 to 12084) Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Ueberback,T., Berry,K., Bass,S., Linber,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,C., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.		

TITLE	Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae A39
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE	20150255
PUBMED	10684935
REFERENCE	2 (bases 1 to 12084)
AUTHORS	Read, T.D., Brumham, R., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Ullrichbeck, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwin, M., Nelson, M., Debou, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
TITLE	Direct Submission
JOURNAL	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
COMMENT	On Jun 1, 2000 this sequence version replaced gi:7190805.
FEATURES	Location/Qualifiers
SOURCE	1..12084
gene	/organism="Chlamydia muridarum"
	/db_xref="taxon:83560"
	/note="synonym: Chlamydia trachomatis Mopn"
	complement(96..713)
	/gene="TC0779"
	complement(96..713)
	/gene="TC0779"
	/note="conserved hypothetical protein; identified by Glimer2; putative"
	/codon_start=1
	/transl_table=1
	/product="conserved hypothetical protein"
	/protein_id="AAF39582.1"
	/db_xref="GI:7190806"
	/translation="MGGLDLKISVGDPSGKTEACQVFEELGAFVISADKVSASF LVPSTSGORVLDLGEPIEYENTLNKALAKVFGSDLLSLKILHGVCFVVEE NPTGVOKKSLFVVEPLLEYGVADMDRVYLISADTIEIKREKRLKKTGSDTSF DLKARFSLSEKILRADYVIEENGTKEFRHKVKKCFALKGI"
	complement(699..3298)
	/gene="TC0780"
	complement(698..3298)
	/gene="TC0780"
	/note="similar to GB:J01663 SP:P00582 GB:J01664 GB:V00317 PID:147312; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=1
	/product="DNA polymerase I"
	/protein_id="AAF39583.1"
	/db_xref="GI:7190807"
	/translation="MKKFLILDASGIFERAFVFLPEMRGNGSTQAVPGFISDKL VNDSPHYMAVFGPPNNKKSROELVADYTSNRRKEDIPEQIALPKOYCELTIC LEKSGVEDDVIAVAKKAAVEGEVICTADKLLQLVNSVSVFPMKEDELTKE VLLOFVAPDQIDADYIALVGDSDNIGVAVGCCGKKQAFLKEFQVEDLVANTORS SKNOMEIOEIEETLLSKRLATHTDMLPDKTEMFSSQTAADVADLNFYLOHGK ALVHAFPTSETSISLOYKNDPSYOMWLKINLKGVEVYCAATGEHLSIRLHGVALS GGDEVFYEISDDEILALKAFAFDATDKFFGYTKRDNALKNHGIDVYATDVILA OHLVNGAKISFOTLLVESGHTQEAATFAKBEWGSLLPVONLPONPAQYGTIVSRIL SIKYLDELKEELKLSIKFNVEDQPLEELIFAMCAMPDSEGLVLADLDSKELET YTOQIVLAGEFENIKSPKLSLVLYHRLGIFKPYDKAKSTAEVLEALDSHEIISKI LAFAPETKMLSTYVRALPROLDLSTHVPHPNPGVGTGRLOCDNINIPVPER GKSLREAFREKENDYFLADYSOVELRPLAHLSOVEDTLRAPDSEDLITPVSQF NVPDLQYTERORYAKAVNGLATGCQATGISKULISVNAOGLIDATFARIPQASE FITHTIEQASKQKVTVMGLGERILTSWESSPGKARASGRVLAFTIRGSAAEILKLA MLNLSRELVSRLGLSRLLQIHDELLFVEKPELEEMKVLVQEMESAMSLVPLVNV VLICKNNAEC"
	complement(3312..4310)
	/gene="TC0781"
	complement(3312..4310)
	/gene="TC0781"
	/note="similar to GP:147868; identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=1
	/product="protease IV; putative"
	/protein_id="AAF39584.1"
	/db_xref="GI:7190808"



OY	425	TGCGACCGTTTACAGGCCATCTCCTACCTCAGGATTTGCAAGACACGCTGGCATCTTAA	484
Db	5712	GCCCGCAAAACCTGACACTATTCTGCTCGAGAGCTCGAGCTTTCGTAGTATGTTTCA	5653
OY	485	AAACTGACATTTGCTGCACTTTATTTACTGCTGCAACTATGGGAGACGCATCTATC	544
Db	5652	GTACTGGAGACAGCTTGTACTACGTAATGTCTGAGCATATGATGACTTCCGTTGTTTATC	5593
OY	545	TCCTAATGTTGGGGATTGCTATATGAATAATACAAAATTCACAGACGAAGCGTTTCTA	604
Db	5592	CACCTTATTTTGGGGGATGCTCAACCAACATTCAGATGTCGTGAAGACGCGCTTTTA	5533
OY	605	CGCTCTTTGGGATCGGACCTAATATTCTTACTAGCTCTTGGCTGCAATGTTTG	664
Db	5532	CGCTCTATTAAATGTTGGATTAAATGTCTCTTATTTGTTGGGAGAAATTTCTCTCG	5473
OY	665	--GCGTTCAAGTTGAGACTTCGCTTCTGAAAGTGAGATCCCTTGGGGAAATTTCTTT	721
Db	5472	GCTGGGTAAACATACGTAATTTCTTCTCTATAGCTGTGGATGCTTGGCACGGGGATTT	5413
OY	722	ACGCTTTTGTATGAGCTATGCACTATTTGCTATCTGACTTGTCTTATAGGCGACTTACTG	781
Db	5412	ACCTAATATTACTCTGTATGTTCTGCTCGAAGAGATTGATCCCTATTATACCGAA	5353
OY	782	GATCAATTAAGACGTATTGACCGCATCTCGCTTCTATTAATCCAGAAATTCAGAAAGG	841
Db	5352	ACTGACCACTTTAAACAGAAAGCGCTCTATTAGAGACGATTTGCTGACGAATGTC	5293
OY	842	GAAMAAAGTGCTAAACCTTAAATGSAATATG-----AAAGATAGCTTCT	886
Db	5292	GGTTGCTCAGCTCAGACAGAGAAAGCGCCCTAAGCAAAAGCAAAAGCTTCTGTC	5233
OY	887	CTATCTGTATAGATCTCTCTTATATTTCTTTATTAATCTCTTGTATTGTGCTATGAT	946
Db	5232	TGTTCTGTCGCTCCGCTCCGCTATTATAGGAATGCTGTGTTGTTGCTTACAACTT	5173
OY	947	TTTGATTAACTTAATGCAATGACTGTTGGAAATGCAAGCTGAACATCTCATATAT	1006
Db	5172	AGCTATCCATCTGTAAACAAATTTTGTGGAAGAGCAAGCTGCTCAAACTATCTCTCG	5113
OY	1007	GAATGACTATAGTAGTTCATGTTGGGAACTTCTCTGACTGAGCGTAGATTCGCTACT	1066
Db	5112	AGTAGAATTTAATTTCTTCACTATGATAGATTAACGTTTCACTGGGATATGTCTGCTCT	5053
OY	1067	TATCATCTATTTTGTGTTGTTGTAAGCTCATCTGTAAATTTGGATGTTAACTGAGCCCT	1126
Db	5052	GGCAGGGGTTTTGCTCGACAGCAAGATATCCGCCGCTGGGGGTGACGTGTTGAGACTT	4993
OY	1127	AGTACCTCTGTCATGTTCTCTTAACAGATATCGTTTCTT--CGCTCTTGTATCTT	1183
Db	4992	GATTAATCCCTTAACGATGTGTAATTAACGGGGGGGCTCTCTTGTGGCATTAATGCTGT	4933
OY	1184	TAGAAACCAAGCTTCTGGGCTGTGCTATGTTGGTACAACCTCCTCAATGCTAAGCTGT	1243
Db	4932	AAAAGGCGATGCGATGATTTTAAAGCGGATTTCTTGATTTTCCCGCATGTTCTACAGC	4873
OY	1244	GGTGTGCGAGCTATACAGAAATATCTTTGCAAAATCCACAAATACGCTCTTGAATC	1303
Db	4872	TTGGCTGGAGGCTGTGCAAGACGCTCTTCTCTAAGCAATTAATTTACTTACTTGAATCA	4813
OY	1304	AACTAAAGAAATGGCTATATCTCCCTCTTGACCAAGAGCAAAAGTCAAAAGTAAAGCTGC	1363
Db	4812	AACTAAAGAAATGGCATTTATCTCTTTGGAAGATGATGAGAAATTTATGTAAGGCGCG	4753
OY	1364	TATTAATGATGTTGGCGCGCGCTTGGGAAATCAGAGGAGGCTTAAATCCAAAGATT	1423
Db	4752	TATTAATGAGAGTATTTCTAGGGTGTGTAAGTGGGGGATCTTTGGTTTACCAAGACTT	4693
OY	1424	GCTGTTATCTGAGAGATTTTGGACTATGACCCCTTATCTTGAGAGATCTCTCTTT	1483
Db	4692	ATTGATCATTTTCTCTCTGTTGTCAGATGTCAGATGAGATGCGATTACTATTTGCTTAATTT	4633
OY	1484	CATCATGCTATTATTGGTTGCTTCTGCAACTAAGTTAAACAAACATTTCTTAGCGCACTC	1543

Db	4632	ACCTTTAGAGGATGCGATTGGTGGTAGTTGGTTGGTTAGCTAAAGATATTCGTTAGAAC	4573
Qy	1544	TGCTCTTAAGAACAAAGTGGCTCAGAGATTCACGCT	1583
Db	4572	CGCTGCTTAGCTAGTAAGCGAGCGGCTGCGTAGACAGCCGCT	4533
RESULT 32	AE008630	10893 bp DNA linear BCT 14-SEP-2001	
LOCUS	AE008630		
DEFINITION	Rickettsia conorii Malish 7, section 62 of 114 of the complete genome.		
ACCESSION	AE008630		
VERSION	AE006914		
KEYWORDS	AE008630.1 GI:15619810		
SOURCE			
ORGANISM	Rickettsia conorii.		
REFERENCE	Rickettsia conorii.		
AUTHORS	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; spotted fever group.		
TITLE	1 (sites)		
JOURNAL	Ogata, H., Audic, S., Barbe, V., Arliguenaue, F., Fournier, P.E.,		
MEDLINE	Raoult, D. and Claverie, J.M.		
PUBMED	Selfish DNA in protein-coding genes of Rickettsia		
REFERENCE	Science 290 (5490), 347-350 (2000)		
AUTHORS	2 (sites)		
TITLE	Ogata, H., Audic, S. and Claverie, J.-M.		
JOURNAL	Selfish DNA and the origin of genes		
MEDLINE	Science 291 (5502), 252-253 (2001)		
PUBMED	3 (bases 1 to 10893)		
REFERENCE	Ogata, H., Audic, S., Renseto-Audiffren, P., Fournier, P.-E., Barbe, V.,		
AUTHORS	Samsom, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and		
TITLE	Samsom, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and		
JOURNAL	Raoult, D.		
MEDLINE	Direct Submission		
PUBMED	Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine,		
REFERENCE	CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean		
AUTHORS	Moulin, Marseille Cedex 05 13385, France		
TITLE	A public version of R. conorii genome database is accessible at		
JOURNAL	http://igs-server.cnrs-mrs.fr. The database intends to provide		
MEDLINE	updated data. Annotation of the genome is an ongoing task whose		
PUBMED	goal is to make the genome sequence more useful. Comments to the		
REFERENCE	authors are appreciated.		
AUTHORS	Location/Qualifiers		
TITLE	1..10893		
JOURNAL	/organism="Rickettsia conorii"		
MEDLINE	/strain="Malish 7"		
PUBMED	/db_xref="taxon:781"		
REFERENCE	complement(326..1579)		
AUTHORS	/gene="ampG1"		
TITLE	/note="RC0718"		
JOURNAL	complement(326..1579)		
MEDLINE	/gene="ampG1"		
PUBMED	/codon_start=1		
REFERENCE	/transl_table=1		
AUTHORS	/product="ampG protein"		
TITLE	/protein_id="AL03256.1"		
JOURNAL	/db_xref="gi:15619811"		
MEDLINE	/translation="MLNRLCIIMLFGLISGFNIMITGNTLANYLAKEDIALOTIGIT		
PUBMED	LSFTLPISINFLAPIFDVAOIKYLNKIFGRSLSCISNALIFLIYFSPDPST		
REFERENCE	NLLAPFALISFSPSAODTISALREIYFKRSIGTSQIYFGRVCMLAGSAA		
AUTHORS	YLSIYTFNEIYKIFAGLVYIYLILIVGICIKYDLBNHIQIKDKINKONKKRI		
TITLE	INFIYNAKPIGSVYFIILILIFLVLRPLNLNVMINPELHLEVDAAEIASVGRF		
JOURNAL	CGVGAIGGLVGVIMKHKNLSIFGIIHAIHGHLIFEFIEINKNSILLETITIG		







QY 79 GAAATTCGCGCTCTTCTTGGCCGATACATCTACGAGCTAAAGAAAGTTCTGCCAA 138  
 Db 2796 AAAAGTTAAAGAAATATATGCGCTATAGAAAGAAATTAAGCTATTTTATACCTA 2855  
 QY 139 TGTTCCTAATGTTCTTGTATATACATTTAACTATAGGTGTGACGATACAAAGACA 198  
 Db 2856 TGGCTTATATGATGTATGATATCCCTTTTAAATTTGGGCTTTAAGATCTATCAAGA- 2913  
 QY 199 CTCCTATTTGGGAGCTCTGCTGTGGGAGAGCAATACCTTTCATCAAGTTTGGC 258  
 Db 2914 - - - - -TAGTTAGTACTACCTCTATAGGCGGCTGAATTAAGCTTTAAATTAATGCT 2969  
 QY 259 TTGTTGTCCTCCGCTCTTATATCTTATGCTTATATGCTTATGCAAGCTAGATATTTAA 318  
 Db 2970 TAGTATTTACCTCCGCTGCTATTTTACCGCTATTTATGTTTAACTTATGTAATTA 3029  
 QY 319 GTAAACAGGCTTATTTTATGAGTGGAGAGCCCTTTTAAATTTCTTGGCCGTTC 378  
 Db 3030 ATTTTGAATATATTTTACATTAATAGTGGCAGTTTCTTACTATTTTCTTAATTAATG 3089  
 QY 379 CGACGTATTTATCCGCTAGCCGATGTTTATACATCCACAGAAATTTGCTGACGTTTAC 438  
 Db 3090 CCTATATTTATTTTCCAAATCAAGATTTTATCATCTCAATGATGAATGAATTAAT 3149  
 QY 439 AGGCCATCTACCTCCAGGATTTGCTAGAGACTGCTGCCATCTTAAAGAACTGACATTTG 498  
 Db 3150 TAATTTGCTTACCTACCTATTTTAAATGCTTATTAATTAATGATGTAATGATGATG 3209  
 QY 499 CTGATTTTATGTAATCTCTGTAAGTATGAGGAGCGTCACTGCTATCTAATGTTCTGG 558  
 Db 3210 CACTGATGTAATTTTCCGGAATATGAGAGTCAATGATCAATTAATGTTTGGC 3269  
 QY 559 GATTTCCTAATGAATTTCAAAAATCCACAGAAAGCAAGCTTTCTACGCTCTTTTCGTA 618  
 Db 3270 AGTTTCAATCAATCTTTTGTACTCTTAAAGCTAAAGATTTATCCCTTCTTGA 3329  
 QY 619 TCGAGCTAATATTTCTTACTAGCTTGTGCTGCAATTTTGGGCTTCAAGTTGA 678  
 Db 3330 TGGTGGTAATATGCTGCTTATATAGCAGGAGCGTACTTCTTTTCTTACACGGGC 3389  
 QY 679 GAGCTTCGTTTGAAGGTAGATC-----CTTGGGGAATTTCTT 720  
 Db 3390 AGGAGCTATGATTCAGAAATTTATACCGATTTTAACTCATCTGCCGAATGCTA 3449  
 QY 721 TACGCTTTTGTGCTATGATATTTATCTGACCTGTTCTTATGCGCACTTCTGCT 780  
 Db 3450 TAATGCTTCAGCAATCATGCTCAATTAATTTACTGCAAGAAATTAATGCTTCTAT 3509  
 QY 781 GGATCAATTAAGAGCTATTTGACCGATCTCGCTTCTATATGCAAGAAATGCAAAAG 840  
 Db 3510 TTAGATATATATGATTTATTTTAAACAGATTCATTAATGTTTATGCAAAAAG 3569  
 QY 841 GGAAGAAAGTGTCTAACCCTAAATGAATATGAAGATAGCTTCTCTATCTTGTATAG 900  
 Db 3570 TTACAGCTAAATGAACAAACCTTGGTAAATGAAGATTAAGTAATTAATCACT 3629  
 QY 901 CTCCTATATCTTTTATTAATCTCTGTTATGCTATGCTATGCTATTTGATTA 960  
 Db 3630 CAAATATATAGCTGTATGCTTATTAATATCTGTTATTAATTAATTAATTAATAG 3689  
 QY 961 TCGAAGTACTTGAAGAGTCAAGTGAATCTCATATCTAATTTGATGACTATAGT 1020  
 Db 3690 TTGAAGCACTTGAAGCAAAATTAAGAAATTTACATCAACACTATAGCTATGTTA 3749  
 QY 1021 AGTTGAGGGAACCTTCTCTTCTGAGCTGCGTATATCCGTAATATCATGATTTG 1080  
 Db 3750 ACTTATGAGGAGCTTATATTTATTTGATGAGGATATCATGCTTACTTCAATGATAA 3808  
 QY 1081 TTGATGTAACCTCATTTGTAATTTTGGATGTTTAACTGAGCACTCATCTCTGCA 1140  
 Db 3809 - - - - -GGAATTAATTTTGTGAAGACCTGCTGCTATTTGCGCATTTAATTAATCA 3866  
 QY 1141 TGGTCTCTCTACAGTATGCTTCTCTGCTCTGTTATCTTTGAAGCAAGCTTCTG 1200

Db 3867 TGTCTCATTTACGCGCTTGATGTTTATCTTATATATTTTATGAGAAATAGAG 3926  
 QY 1201 GGCTGCTGCTATGTTTGGTACACTCCTTCATCTAGCTGTGTTGCGAGCTATAC 1260  
 Db 3927 AATGTTGCTGATTTT-----AATCTCTATATGACGCAATATTTGAGCAATTC 3980  
 QY 1261 AGAATATCTTTGCAATCCAAATATAGCTCTTTGACATCACTAAAGAAATGGCT 1320  
 Db 3981 AGAATATCTGATTAATCATCAATATCTCTATTTGATTTCAACAAAGAAATGGCT 4040  
 QY 1321 ATATCCCTCTTACCAAGCAAAAGTAAAGTAAAGCTCTATGATGATGTTGCG 1380  
 Db 4041 ATATTCCTTTATCTTATAGAGCTTAACTAAGGAAAGCTCCGTAGAGTAAATAGTA 4100  
 QY 1381 CCGGCTTGGAAATCAGAGAGGCTTAAATCA-----ACAGGTTTGTCTTATCT 1434  
 Db 4101 CTAATTTGCTTAATTCGTTGGAGATTTATACGCTTTGATTTATTTATTTCCGA 4160  
 QY 1435 GTGAAGTATTTGAGCTATGACCCCTTATCTGCACTGATCTCTTTTCAATGCTA 1494  
 Db 4161 CTGCTACCTTTGATTTCTATTAATATATTTATTAATTAATTTATGTTATGATGAGTT 4220  
 QY 1495 TTTGGTGGTTTCTGCACTAAGTAAACAACTATTTCTAGCGAGTCT 1544  
 Db 4221 TATGATTTTGGACGTTAATTAATGAAATATGATATGAGCTGTCT 4270  
 RESULT 33  
 AE008626  
 LOCUS  
 DEFINITION  
 Rickettsia conorii Malish 7, section 58 of 114 of the complete genome.  
 AE008626 AE006914  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Rickettsia conorii.  
 Rickettsia conorii.  
 Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia: spotted fever group.  
 1 (sites)  
 Ogata, H., Audic, S., Barbe, V., Arliguenave, F., Fournier, P.-E., Raoult, D. and Claverie, J.-M.  
 Selfish DNA in protein-coding genes of Rickettsia  
 Science 290 (5490), 347-350 (2000)  
 2 (sites)  
 Ogata, H., Audic, S. and Claverie, J.-M.  
 Selfish DNA and the origin of genes  
 Science 291 (5502), 252-253 (2001)  
 3 (bases 1 to 10200)  
 Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.  
 Mechanisms of evolution in Rickettsia conorii and R. prowazekii  
 Science 293 (5537), 2093-2098 (2001)  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 4 (bases 1 to 10200)  
 Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. and Raoult, D.  
 Direct Submission  
 Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine, CNRS UMR 6020, Universite de la Mediterranee, 27 boulevard Jean Moulin, Marseille Cedex 05 13385, France  
 A public version of R. conorii genome database is accessible at http://igs-server.cnr-mrs.fr/. The database intends to provide updated data. Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful. Comments to the authors are appreciated.  
 FEATURES  
 location/Qualifiers

source 1. .10200  
/organism="Rickettsia conorii"  
/strain="Malish 7"  
/db\_xref="taxon:781"  
73. .1608  
/gene="tlc4"  
/note="RC0666"  
73. .1608  
CDS  
/codon\_start=1  
/transl\_table=11  
/product="ADP ATP carrier protein"  
/protein\_id="AAL03204.1"  
/db\_xref="GI:15619753"  
/translation="MTINPSNVNSSKINSYPSKLDYITWPIKRHEVSKFLFTLLM  
FCILFIONLRALKDSIVTWIGAEIISFLKFMGMPASFLMTAIVKLVKMAENI  
FYLIISIFLFFELFAVYIPFNHEMLHSPYVQNLMAFLNKLKFWIIMLSKMSFLF  
YIAELMNVVEFALLFWOFVNNITTVESKRFYPLFGILISQGYLAGOPLNLSNIN  
DYVTKRALOSSPHLISQIILITVLVLGIIATVPMLNKHVLDKHEMALFRKAK  
KSMTASFQMLSSRHILRLATLTCYGAIINLVEGPKAAATIKYKTPETAAGIG  
SYLSTGVETFLFVVLGSNIVRIGMTAAVTPPLIVEITGILFFAVNNERFAGLI  
ANFLTDPALAIITGAIQWVLSKSSYTFLEDSIKEMAYVPLDEIKIKGAADVIG  
TKLGSSGSAFLOSIVFILPSASYSISTCMIIFITLCLTWMATKALNKKEYNSIK  
ESQ"  
repeat\_region 1691. .1724  
/note="REP02, RS3-like repeat"  
repeat\_region complement(1725. .1749)  
/note="REP02, RS3-like repeat"  
repeat\_region 1781. .1810  
/note="REP02, RS3-like repeat"  
1889. .4969  
/gene="sca4"  
/note="RC0667"  
1889. .4969  
CDS  
/gene="sca4"  
/codon\_start=1  
/transl\_table=11  
/product="cell surface antigen"  
/protein\_id="AAL03205.1"  
/db\_xref="GI:15619756"  
/translation="MRGFMSKDGNDLTSEFDPLANKYEYEEKOTLEBQKEFLSOT  
TPALEADGFIIVSASPAQSPTMSALSNGISPSQSPDITKAVRETTIOPKDNLI  
EQLIKDLAATITDRDLAQKREIEEKEKRTISTFGNPNRREFIDKALENBEIKK  
LESLIAGIKNVNHTSASGYPGEFKPVQWENNVSANDLRATVKNDADECTLE  
TVYTKRPTLAKODGTQVQISYREIDEPILDKADGSMHLSMVALKADGTKCKDA  
VYFAHYEGBPCKPOLKEISPRPLKFACTGDALAVIEHGGEIYTLAVTRGVKEM  
MKEVELNQGQSVLSDAEDIIIGQSGSEQPLITPOQTSSVEPQYKQOAVPIPT  
NOPLOPESQMPQSOQVNPMLNATLALSGMODLINVNAGLTKAIDSKODILKE  
AATAILNNEKSDIAEKQANITIALAENVNNKNIKPKPAKYCVNAVLETTKNDONTPL  
EKSKMEATYAIYLSNENLEPPKQOOLEKAVDGLSKDDASBRAKIDGKIDVYIKS  
NLSPEDEMLIAGVDKVNSELNAEKOKLGLSVLKRGVEQVLSPAQOOLMOQHLVKI  
MABOTKDTIKRVNDILFDPLSNTELTNTNIOAITSNVLDGPATAEYKGIIOAITWT  
IAGSSLEAODKAAIINGVETIATHSPTSLSPNKALIMASAEKGIASEQNLNPLREL  
MTGGLVDGIYEKGGEPEITKAVSSGIDNANVINSEKALKAKDAASEALDPTDOL  
TEGPKQNIIEHKPHDDIYNKAREVITANVPVIEALEKSEPYSAEERTYQOQSTL  
NITSKLAVEKVNPRAMLSPNGMLKTLIEKKEEIKVDELKVAFGTKSSTEEQOST  
KTNLIDKTLKSEVRLOTIDKLQEQKRSEAIINPSYKTEDVAVVSGSKSLKISKN  
PDIEKAKMVGRDVRNIKNIKIMGALMNARDIIQSENLNKSTPIKRESSPPQR"  
gene /gene="RC0668"  
complement(5088. .5465)  
CDS  
/gene="RC0668"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAL03206.1"  
/db\_xref="GI:15619757"  
/translation="MGRKNDIQKIDSFTGKITYSLRLAKLSRQOLAEDIVDTHOOL  
QKYEKAINRISVRLVLAELDRNIDYFPEGLEENKPPQVTHQHRMCIEVSNNEM  
KINSTEEOQAVNNLVKLAAGKN"  
gene complement(5533. .5691)  
/gene="RC0669"  
complement(5533. .5691)  
CDS

gene /gene="RC0669"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAL03207.1"  
/db\_xref="GI:15619758"  
/translation="MFFNNYLPNDYKRLKRCYIVAVYNILMWYFYLLDLKFNIGIKLL  
ITTCFII"  
5987. .6814  
/gene="RC0670"  
5987. .6814  
CDS  
/gene="RC0670"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAL03208.1"  
/db\_xref="GI:15619759"  
/translation="MKRIFNCLVYALFRTIEDGVENSGYMSPHILSTPPLVELLA  
LTSFLGASELGQNFQIFLESLEQAVESIEKRIRELLSAPPOSIMNLAVIGSIWAS  
SPECLRTIINRWYQIKSPPYIRRLSLSTIOPILSALITTFMFLVLPILFTKIP  
IILEETIKRYIILNFIYFLILILFLGASSLVYILPNVKLNFDPGALLVVIIMI  
ISGLISTYIVVYQNLMLMGSGSIIVTLIFFYIIMFIYGAEFNVLKKNENIE"  
complement(6964. .7797)  
gene /gene="RC0671"  
complement(6964. .7797)  
CDS  
/gene="RC0671"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAL03209.1"  
/db\_xref="GI:15619760"  
/translation="MKISHINYSFINTLSGRLOKNKOLEKIIIAQFHODYLIPIIGV  
EIEFYLSHNIDIAKFEILSRKYLARFKIAIKKEKGNQOEIDPSPVNILOYIKNL  
EVTKTIKKMAQOLNGYIDFSPKPLDYGSMHPHINENSEFNDYYITIAAGDLCYHM  
LDPLIAFMTPLDYARINKKEMAPTHISYCGNRSVAVRPNAPKRLERLSSPED  
PYIAIFTIKSIILALKSPSSLOTIEKITGNAPDQYNLPLPFSOASFWLKRPF  
K"  
7796. .8518  
/gene="RC0672"  
7796. .8518  
CDS  
/gene="RC0672"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAL03210.1"  
/db\_xref="GI:15619761"  
/translation="MEGLINSVYYKIFDKFNNSHRYIKKNSINERPIVENKSI  
TYFKKODIILNVOHGNOTVNADESTIAVPEADGSTTKNKLIAVQSGALCVPIILA  
SGGKITGAHACHKGSINNINISNYTKITTEKAKNLIAVIGALIAOSSYEVDEYK  
AFSLIDNNKQFPIHSHIKENHYFDEPAFVELKLEKXGVKDIRINIAEDITYTNPKYPS  
KRSYHLOEYPNONILSAIVMK"  
8597. .8630  
/note="REP01, repeated element"  
8622. .8786  
/gene="RC0673"  
8622. .8786  
CDS  
/gene="RC0673"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAL03211.1"  
/db\_xref="GI:15619762"  
/translation="MAIQIKIKFKCLEFTTGILHQLLCNFLVMTGKPYATMLARNN  
DIVLYAMOK"  
8650. .8687  
/note="REP01, repeated element"  
8811. .9362  
/gene="def2"  
/note="RC0674"  
8811. .9362  
CDS  
/gene="def2"  
/codon\_start=1  
/transl\_table=11  
repeat\_region  
gene  
CDS

```
/product="polypeptide deformylase [EC:3.5.1.31]"
/protein_id="AAL03212.1"
/db_xref="GI:15619763"
/translation="MNODKPYQIVYAPNDIEFKQAEYIDVDNIRIVDKMLONLH
IERAVLGANWVGLIKRIAVLDLHNNKSSPITVINPITFSEKQFIEGSLSPFG
IEASITRSKAIKRYLDINGNKOELCAAGFLATYIOHEIYLNKGTFLDLSKLKRDY
LLKMLKHILKPHIHSGCRH"
gene
complement(9295..9522)
/gene="RC0675"
CDS
complement(9295..9522)
/gene="RC0675"

Query Match      13.2%; Score 215.8; DB 1; Length 10200;
Best Local Similarity 49.7%; Pred. No. 5e-41;
Matches 668; Conservative 0; Mismatches 662; Indels 15; Gaps 4;

OY  74 TTTTGGAANAATTGGCCTCTTCTTGGCCGATACATCTCAGAGCTAAGAAAGTCT 133
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  126 TTTTCTAAACTTACCGATTATATCTGGCCATTAAGCCATGAAGTTTCTAAATTTT 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  134 GCCAATGTTCCATAGTCTCTCTGATATACATTACATAGCGTGTACGGGATACAAA 193
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  186 ATTCATCACTTTATTAATGTTCTGATATTTATTCACAAATCTAATCAGAGCTTTAAA 245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  194 AGACACTCTTATTTGGGAGCTCTGTTCTGTGACAGAGCAATACCTTTCATCAAGTT 253
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  246 AGATAGTATTTGTTA-----CTACTATGATCGGTGCCGAGATATCTCGTTTGAAGTT 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  254 TTGGCTGTGTTCCCTCTGCTATTTATCTTATGCTTATTTATGCAAGCTAAGTAATAT 313
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  300 TTGGGGAGTGAAGCCGTAGCTTTTATGACTGCTATATATGTAAGAGCTTTAATATA 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  314 TTTAAGTAAGCGGCTATTTATGACAGTGGGAAGCCCTTTAATTTCTTGGCCCT 373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  360 GATGAAGACAGAAAATATATTTTATCTTATATATCATATATTTTAACTTCTGCTCT 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  374 GTTCCCGACTGTATATTTCCGCTACGCGATTTTACATCCACAGAAATTTGCTGACCG 433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  420 TTTTGCTACGTTATTTTCCAAATCATGAATGCTGCTTTCATCCTCGTAAACGTTTA 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  434 TTTACAGGCACTCACTCCAGATTTGCTAGAGCTGTTGCCATCTTAAGAACTGAGC 493
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  480 AATTTAATGCAAGTTTACCTAATTTAAATGTTTATATGCTTTTATCAAAATGAGAG 539
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  494 ATTTCGTGATTTTATGATCTGCTGACATATGGGAACCGCATGCTCTATATGTT 553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  540 TTTTTCGTTATTTTATATATAGCCGAATATAGGCCAAAGTATTTGCAATTTCTTTT 599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  554 CTGGGATTTGCTAATGAATATACAAAATTCACAGCAAGCAAGCGTTTCTAGCCTTTT 613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  600 TTGGCAGTTTGTATTAATATATCTAGACATCTAAGAAATTTTATCCGTTATTT 659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  614 CGGTATCGGAGCTAATATTTCTTACTAGCTCTGTGTCGCAATTTGTTGGCCTTCAAA 673
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  660 CGGTTTACTTACTCAAAACGATATTTATTTAGCAGGCGAGTTTAAAGAAATCTAAGTAA 719
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  674 GTTGAGAGCTGCGCTTTCGAGAGGTGTATCTTGGGGAATTTCTTTCGCTTTTGAAT 733
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  720 TATTAATGATTAAGTCACTAA--TAAATTTGCAATTTGCAATGCTTTTCAATCACTTTC 776
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  734 GGCATGACTATTTGATCTGACTTGTCTTATGCGCACTACTGCTGATCAATAAGAA 793
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  777 TATCAAAATTTATCTAATATAGTATTAATTTTAGGAATATATAGCATTTAAACCTTTTGG 836
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  794 CGTATGACGAGCTCTGCTCTATATATCAGAAAGAAATGCAAAAGGGGAAAAAGGTGC 853
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  837 GTTGCTTAATCATAAAGTACTAGACAAAGAGCATATAGCGCTTACAGGTTTAAAGCAAA 896
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  854 TTAACCTAAATGATATGAAAGATAGCTTCTCTATCTTGTAGATCTCTCTATATTTCT 913
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  897 GAAAAAATCTATGACTATTTGCGGAAAGTTTTCAGATGCTTCTATGCTCAAGACATATTAG 956
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  914 TTTATTAATCTCTTCTGTTATGCTATGATATTGTCATTAATGAAAGTGAAGTCTTG 973
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db  957 ATTAATTTGCAACTTTGCTATCTGATGAGCAATTTGCAATATTTAGTAGAAGTCCCTTG 1016
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  974 GAAAAGTCAAGCTGAAACAGTCAATATCTATATATGATGACTATAGTACCTATGAGGCA 1033
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1017 GAAACAGCAGCAGCTACTAAATTTATTAACCTCAACCGAATATGACAGCTTTTATAGGAAG 1076
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  1034 CTTCCTCTCTGAGCTGGCGGTATGCTGCTATTTATGATGATTTGTTGGTGAAGCT 1093
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1077 TTAATTTGAGCTATAC---CGAGATATTTACTATTTTATTTGCTGACTGTTGTTCAATAT 1133
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  1094 CATTCGTAATTTTGGATGTTGATCTGAGAGCCCTAGTCACTCCTGATGATGTTCTCTTAAC 1153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1134 ACTTAGAAGACTGCTGCTGTTTACGGCGCTGTATACACCCCTTAATATAGTTTATATAC 1193
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  1154 AGCTATCGTTTCTTCTGCTCTGTTATCTTTTGAANACCAAGCTTTCGGCTGCTAT 1213
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1194 CGGTATATTTATTTTGGCGTTAATATTTTGAAGATTTGCGGCTTAATATAGCAAA 1253
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  1214 GTTGGGTACAAAC---TCCCTCATGCTGCTGCTGTTGTTGGAGGCAATATACAAATATGCT 1270
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1254 TTTTATTTCTAACCGATCTGCTTTATTTGCTATTAACATATAGTGTCTTCAAAATGCTCT 1313
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  1271 TTGGAATTCACAAATATAGCTCTCTTGTGACTCAACTAAGAAATGCGCTATATCTCT 1330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1314 TGTAAATCAACCAATATATACCTTATTTGATTCACAAAGAAATGCGCTATGTTCTTT 1373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  1331 TGACCAAGAGCAAAAAGTCAAGTAGAGCTCTATTTGATGTAGTTCGCCGCCCTTGG 1390
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1374 AGATCCGGAAATTAATAATTAAGCAAAAGCTCTGCGACGTGATAGGTACAAACTCGG 1433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY  1391 AAAATCAGAGGAGCTTTAATCCAA 1415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  1434 TAAATCTGCTAGTGCATTTTTACAA 1458
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 34
RPXX02      312430 bp      DNA      linear      BCT 12-NOV-1998
LOCUS       RPXX02/c
DEFINITION  Rickettsia prowazekii strain Madrid E, complete genome; segment
2/4.
ACCESSION   AJ235271
VERSION     AJ235271.1 GI:3868717
KEYWORDS    Complete genome.
SOURCE      Rickettsia prowazekii.
ORGANISM    Rickettsia prowazekii; alpha subdivision; Rickettsiales;
Bacteria; Proteobacteria; Rickettsiae; Rickettsia; typhus group.
REFERENCE   1 (bases 1 to 312430)
AUTHORS     Anderson,S.G., Zomorodipour,A., Anderson,J.O.,
            Sichteritz-Ponten,T., Alsmark,O.C., Podowski,R.M., Naslund,A.K.,
            Eriksson,A.S., Winkler,H.H. and Kurland,C.G.
            The genome sequence of Rickettsia prowazekii and the origin of
            mitochondria
JOURNAL     Nature 396 (6707), 133-140 (1998)
MEDLINE     99039499
PUBMED      9823893
REFERENCE   2 (bases 1 to 312430)
AUTHORS     Anderson,S.G.E.
TITLE       Direct Submission
JOURNAL     Submitted (11-NOV-1998) S.G.E. Andersson,
            Siv Andersson@molbio.uu.se, Dept. of Molecular Biology, University
            of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
            On Nov 13, 1998 this sequence version replaced gi:3860788.
COMMENT     FEATURES
            location/Qualifiers
            1..312430
            /organism="Rickettsia prowazekii"
            /strain="Madrid E"
            /db_xref="taxon:782"
            76..1830
            /gene="RP226"
            76..1830
            /gene="RP226"
CDS
            gene
            76..1830
            /gene="RP226"
```

/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA14689.1"  
/db\_xref="GI:3860789"  
/translation="MLNKLCDILFLINLLVTVGYASPPPLPSPSLPIIIVDTDKNI  
STNSIISFEKQFQSKQKKNISSQEQEQQKAIHQESQOIDSREINENQSEPTI  
DFGSTITPSPVSNYIDSKAEXENSTNLAASTYQTDIOYKQDFPSEASBPIDIGNK  
FTSATNHEMKYEAASNDKENTNLSTNITTPWVSPVISIPTADVNYYVPSQSOVQI  
KPTNLTSIRNPIDPLNHHTDLNKEKNESTISNMTTIPTNVSPSIDDTIOTTLNIT  
VPTAEKTSVQTSYVMSHNSHQAOPITPISINTPVETSPTVLRATESVPTINNSOEIF  
VSESESTKQDWTPIPMVLYVDPKKSQSLAEQKNNQDITINNOAESVSSSNV  
TIQKONQVNNATSESTKEFYKNETQMLFLPDDIVLQKLTEDATLEQMDHGYTKLF  
QKREEMIANAEKRLVSEFIKYDDINKNDIYANLSYSAVDNAFRAVENNNLFEELR  
ALDDVYPLLOAKNSTGETLTLSIYNGNYYLAKELVIRGITSVLKNDCEKYPDIAL  
AKGNTNICYMLKAKGYN"  
2025. .4013  
/gene="Rp227"  
2025. .4013  
/gene="Rp227"  
/codon\_start=1  
/transl\_table=11  
/product="DNA GYRASE SUBUNIT B (gyrB)"  
/protein\_id="CA14690.1"  
/db\_xref="GI:3860790"  
/translation="MSDLFSLNKKKKNIVYTNYSANDIEVLDGLEPYKRRPGMYIGT  
TDSNAMHLVSEVLDNADNAVAGFASIIIMIKMDHSITTFDNGRGIPIDNHRFPD  
KSALEVLITLTHSGKFSNNYHTSGGLHGVIGSVNALSKEFIKYKQKGLQSYQ  
SKGAKLIDLSAEASKRLRGTSINFTDPPETFEFKLHPNPKKITYEIASRKYLYRGS  
IEMCEVEYPSDIPKKALINPNPKLDKLSKISLDNLVPEIFSGNIESIVDIKLE  
WALCQNNDSISAFMOSYCNVPTPQGGHESGLSALLRGLKATSEMIGNKANLVI  
EDILETSIVISIFIVERFQGTQEKIVSGVSKLVENITKDHDFHLSDDKVLATH  
LSEHVAIAEERRRKNKERNISRSKSVQKLRPLGKLADCTRTSAEGTELFVEGDSAG  
GSAQARNRETQAVPLMGVLANSTLEKTIINNOAIODEIALACGSJKNYKENL  
RYEKTIIMTDADVDAKHAASLMTPEPFLRMPKLYVEGHLVYAKRPLRYLQSNKITYA  
CDEEEKIKITYKLSKASKAKTEVGRFGLGSMAPQLEKTTMHPKRSLLKVTLEDVO  
NVDKITVDDLMGKPEKRRPQITFEQALVAMDITINKLDI"  
4122. .5447  
/gene="Rp228"  
4122. .5447  
/gene="Rp228"  
/codon\_start=1  
/transl\_table=11  
/product="Tail-SPECIFIC PROTEASE PRECURSOR (ctp)"  
/protein\_id="CA14691.1"  
/db\_xref="GI:3860791"  
/translation="MLLRFIALFLSINCAEGEKETENKTSINOFAVKOPQDVEIRIK  
DYVQVDPROMIDPAINGMLNSLDPHSNYVTDLEDLIDFTTPKGEFGIGVEINYDSG  
AIKTISSIDLPARKAGIKGSDIYGVNDELVSTLGRKATKEKRGTPGTVRLITIK  
EEBAKPOLELTREIVIKIPYKHALEKNNTAVIRITTFNESTISELKAAYKRLTESK  
DNLKGILLDLNNGGILDOAIVASDYFIDSGVITFGRTSSNSETKANFEFLKAP  
KVPWIVLINGSASASEIVAGALODKRAIILGTSFGKSGVCALOTINSAARKLTI  
SKYVTPSGRSIOAGIEPDLIPAKVEYEPVKIDKRFSSSLKNVYLKNMAKNKDS  
NKEKTKNNKQSESELSEIYKKQYAFARADVITGLINTNLETOGAK"  
5653. .7146  
/gene="Rp229"  
5653. .7146  
/gene="Rp229"  
/codon\_start=1  
/transl\_table=11  
/product="HISTIDINE KINASE SENSOR PROTEIN (bara)"  
/protein\_id="CA14692.1"  
/db\_xref="GI:3860792"  
/translation="MHNKQIVRLSFTILFLVYINMLFEYFMIKEMIKOVALEHT  
KIYELDYDINMINKNVISKLHKEDYKLLDODQFINFKITTAQMPFNININISLYLT  
KGNKEITSNMLHMSYDNYKDDSLIEIVITKIDKILKSFSTSAAPLRDAKGTSHL  
LPKVINENSDLTIEHASFVSYTIPVDHNDLPVDAVEFINNTISQNNITSLEAK  
VFTEFIITFIETITISNTNVAQOIIIEQDLETNRNLKAOIEKERTSSNTAFEFAN  
SHELRTPLNALIIGFSEILMSERDEKSNYIKDIDHKGHLISMINDILDSAFASAK  
LKVDNIDLNLKLISSELILIKPRAYAEVELLSRLKEHVHVINADPRKROYLNLNL  
SNVAKTPSGSVATISLEKDELAKLYIKVINDIGIEEKDIPKTLISAFQIDSELSR  
KYEGTGIGLPLTKKVLVLMNGKPDLOSKINKGTITVITTKYDGSIEI"  
7374. .8126

CDS  
/gene="Rp230"  
7374. .8126  
/gene="Rp230"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA14693.1"  
/db\_xref="GI:3860793"  
/translation="MKLITLFLSFLSFGSKSEMPQGRFLKYAVVNNDFENRLDEOE  
QEIRILGVEVLIQHKIDILKONLMINQENETVLEETDDEKRODIFDIALLEGMDH  
VSKRTFENVNDIARYKQAYDALAYKDNKTEKDKFKNITQYRPNNSLISNAYFY  
AEQFKQKRDVNGAINTLKCYKESPKAKSSDGLKLALSLIGELKKQAEACNIIAKLD  
KEPFINRTSYSKMTEDAKIKFGCKINKNKMI"  
8123. .8785  
/gene="Rp231"  
8123. .8785  
/gene="Rp231"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA14694.1"  
/db\_xref="GI:3860794"  
/translation="MTDILDEVLSQDNEEKRLIFPKLVPIIIISIVITIMYITIN  
NKDNRIKNNQKNDILVKTIGLDTTKNKAIAFTLENLTSSMTKIKETAAEQVAI  
RMSERKYLGAKNLNLNKIIDNEKSEISTAAARIAMCIVLDDQSLNIDQCKLOKYL  
YFDDERKPFMTATITIKAIIDIKHNMKTQAEKNLKNLASNNVSDLLKDAQKALLVSL  
SK"  
8812. .10059  
/gene="Rp232"  
8812. .10059  
/gene="Rp232"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="CA14695.1"  
/db\_xref="GI:3860795"  
/translation="MKRKLALLLPILISNGIGSKSVKNIVYDLPKLYIQTNEPY  
LDSNTKIPFNVMNLKKNQYSPASKMIAEPVIGDMITVLDIIRANISAEIKNKII  
WSYLSKHKNDNTIGGSLHNGKLYITYCARLLIVYDAKSGYEIIKELPDIIRIP  
IALNDHTLIQVOTISNOTIALDSELTAKLMDHESIATLSTSYSMIPVQHDNVIVYTN  
TGOVLATINNGEYKKNPEFTNLNDHAIINFPDSSILCTPYHDSMLYIATGCGIKI  
KNLITGSVMQINADDIQSLSIGNSLFIINNAQIALNPPETGKVFADVETEN  
DKRLKSTFPLVEYGVNNDORSNLVSDGILYFNVIDSNGLKNMHHIIKIIKNIRY  
YGLRSNTPLVEFYSTDQVIFGSOI"  
10279. .10746  
/gene="Rp233"  
10279. .10746  
/gene="Rp233"  
/codon\_start=1  
/transl\_table=11  
/product="50S RIBOSOMAL PROTEIN L13 (rplM)"  
/protein\_id="CA14696.1"  
/db\_xref="GI:3860796"  
/translation="MKTIYSAKPSPIETKKMYVIDAKNVYLCRLASRAVIMRGHKKSF  
TPHDCGDNITIIINAEHILKTLGKLNPKDGVYRHTGFPFGIKDTPAGILSGKYPE  
RVIKMAVKRMITRNVLGAKOMSNLYVANCHEPHMAQOPTIYDFASEPNKK"  
10753. .11238  
/gene="Rp234"  
10753. .11238  
/gene="Rp234"  
/codon\_start=1  
/transl\_table=11  
/product="30S RIBOSOMAL PROTEIN S9 (rpsI)"  
/protein\_id="CA14697.1"  
/db\_xref="GI:3860797"  
/translation="MTELAKTEKVVAKOLTKRESLSYKALTPKAKIDSVSFKFYATGRK  
NATARWLVKVGKRIYVNNKILNQVPESEYVATTILOPILTLTFTDQYDVICTVKGKG  
ISOQKAILHGISIKALDPSAPCEHAIIIRGGLTLRDSRVERKKYQCARKKKTQFSK  
R"  
11279. .11352  
/gene="trna Met (CAT)"  
11279. .11352  
/gene="trna Met (CAT)"  
/gene="trna Met (CAT)"

gene





Db	5691	TGGGACAGATGGAGTTGGTATTATTATTATTATTATGAGAGAGTTATGGCCGTTTAACTC	56330
Oy	540	CTATCTCTAAATGTTTGGGGATTTGCTTAATGAAATTACAAAATCCAGCAAGCAAAAGCT	599
Db	5631	TTTACTCTTTTAAATATGCGAGCTTGCAGAAATAAATCTAAAGTGCAGAAAGACACAGAA	5572
Oy	600	TTTCAGCCCTCTTTGGGTATCGAGGTAAATATTTCTTCTAGTCTGCTGTCGTCATCT	659
Db	5571	TTTTACTCATTTTTTTTACTTTATTTCGACAACTAATTTACTATCTCAGAAAGCCCTAAT	5512
Oy	660	GTTTGGGCTTCAAG-----TTGAGAGCTTCGGTTTCGAAAGGTAGATCT	707
Db	5511	ATTATATTTCTGAAGAGTGCAGATTTCTTATTCCTTTATTTCTTCATCTAAACACTACA	5452
Oy	708	TGGGGAATTTCTTAACTCTTTTGATGGCATGACATATGATCTGAGACTTGTCTTATG	767
Db	5451	AATGAATATCTTTTAAATATCATTTATTATTCAGTTATTTAAATATCCGATTAATTTGTCTA	5392
Oy	768	GCCAGTACGATGGATCATTAAGAACGTATGACCGATCTCGCTTATATATCCAGAA	827
Db	5391	GCTCTCATTAACCTTAATGTATA-----ATCAGTCGTAGAAGCTGACAAAAATAT	5342
Oy	828	GAAATGCAAAAAGGGAAAAAGTGCTAAACCTAAATTAATATGAAGATATGCTTCTC	887
Db	5341	TAAATTTAAAAAACCCAAAGATG--GATATATTTAAAAATTAAGCTTGCGTACAGCTCAAAA	5284
Oy	888	TATCTGATGATCTCCTCTTATCTTTATTTATTAACCTCTTGATTTAGCTATGATGAT	947
Db	5283	GTAATATTAACTCTAGATATCTCGGTTTTATTTTGCTCTTGATGATGCTATCTATG	5222
Oy	948	TGCATTACTTTAATCGAATGACTTGGAAAGTCAGCTCAAACTGCATATCTCTAATATG	1007
Db	5223	AGTATTTGCCCTAATAGAAAGATTATGAGTGTCCAAAAGTAAAGCAACTATATCCGCTACA	5164

Db 5163 AACGATTTTATAGCCTATCGCGTAAGTGTTTTGGACGGGAATCTCACGTTAGTT 5104

[illegible]



ORGANISM  
Rickettsia prowazekii  
Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.  
REFERENCE  
1 (bases 1 to 279110)  
AUTHORS  
Andersson, S.G., Zomorodipour, A., Andersson, J.O.,  
Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K.,  
Eriksson, A.S., Winkler, H.H. and Kurland, C.G.  
The genome sequence of Rickettsia prowazekii and the origin of  
mitochondria  
JOURNAL  
Nature 396 (6707), 133-140 (1998)  
MEDLINE  
99039499  
PUBMED  
9823893  
REFERENCE  
2 (bases 1 to 279110)  
AUTHORS  
Andersson, S.G.E.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (11-NOV-1998) S.G.E. Andersson,  
Siv Andersson, Molbio. un. se, Dept. of Molecular Biology, University  
of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN  
FEATURES  
source  
1. 279110  
/organism="Rickettsia prowazekii"  
/strain="Madrid E"  
/db\_xref="taxon:762"  
63. .1046  
/gene="RP479"  
63. .1046  
/gene="RP479"  
/codon\_start=1  
/transl\_table=1  
/product="OCTAPRENYL-DIPHOSPHATE SYNTHASE (ispb)"  
/protein\_id="CA14934.1"  
/db\_xref="GI:3861034"  
/translation="MNITVYKIQNLKDEVTOLNDLITSLKSDALLKRYGVLYEAG  
GKRIRPLITITAKMEDYKGNHNIKLASAVERITHAATLLHDVYVNDNLRPRKPNV  
IMGSTSLVGDLEFESQSKLMVAGSCIKAMVYLAKASYISEGVOLVTKNERIIT  
TIDEYQIVKSTAEFLGACVEGATIEQVDRSKDVQNGRLGITFEQVYIDDLDT  
LGSKQVGNKNGIDFLGKVTPLPLFLVHKLEQDKQLNMLKSKDKRTKDFVYKRD  
LMLKHAIVETVNYLSLENEANMLNKNIPQNTIKYVLYFSIIRFLYRSY"  
1065. .1138  
/gene="tRNA Arg (CCG)"  
1065. .1138  
/gene="tRNA Arg (CCG)"  
/complement(1194. .1265)  
/gene="tRNA Gln (TTG)"  
/complement(1194. .1265)  
/gene="tRNA Gln (TTG)"  
/complement(1316. .3091)  
/gene="RP482"  
/complement(1316. .3091)  
/gene="RP482"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="CA14935.1"  
/db\_xref="GI:3861035"  
/translation="MRHDIRNLRLKFLTEVNIIEGYIIPNDKYNNEVPEYAKRLEYI  
TGTGSSGSIATICDAEFAFTDGRYLLEQANKKELDLAFKIVDLKEIYKTLDRNKIGY  
DPLFTYOVLANLINNFKNINENLVOKIWMYKPLEPNSKYVLAHDIKFAGVSHNKRIRK  
CRETITSSSVTGCNKNNDDILVILDSASIKNLRASDVNTPLMAFAVITLSTK  
LYLFINPFRIDETIINARPEITIIPEKEFENILRSKRNRILIDSITSVHIMDLTANK  
KVKTIVEPCLAKACKNDIEIKHAIDFIKDAVALCEFAEFYHSSENVNSCFHSH  
EITIEHSGLCLTLQARAKQEGVSDSPFAICGFQENSAIITHYRANPKYAKIEGCGIL  
LIDSGAQYKAGATTDTTRTIIIVGIPCEQKRYTOYVKGHIALTRAKPKNVLTGCLD  
ILAROYLQMDMIDPDHGTGHGVSFLSVHEGPOSILNSKTIIRAGMTLSNDEPGYIP  
GKYGRIENLIVYKNGMLEFEETLSLVPRYASKLIDVALLINDELINIKYEYNNIRAK  
IYNLSTOAKTWNLEINFLQSL"

/protein\_id="CA14936.1"  
/db\_xref="GI:3861036"  
/translation="MSOKLSFMAVALVYSGISVFLPLSLAPFGVSYMGWILS  
LFGAMSLAVBSCLCTKRPKGTGGPHVYVRASFGDPIAFCTGTHYIISFVSTSYVIS  
ALGVLTPFKSOTLIDLILQILALAIILNLKNGKIAKGVEFTTLTKFVPLVGL  
AALFHEINDNTVIAKEVNEFTPTIMGVALLTWEFGIGICATTTACTVADPATIP  
RAIIITGTCVAFVLYIINSIGIIGLIPASELINSKAPYADATLLFGTWSKRVIVIAS  
VICIGTLMNAVLTSGOIALGLAEDGLPEKFAKKNNAPTYGIIISGLSTPLLP  
SNNNPAKQITQIDPSVATPLFVYIGSLAFKVFSSKKNFSYVLLVAIISIIFCT  
WVIYKTPPETILNASTFTIIGLPIYVYGMRLNRIKL"  
complement(4633. .4965)  
/gene="RP484"  
complement(4633. .4965)  
/gene="RP484"  
/codon\_start=1  
/transl\_table=1  
/product="HESB PROTEIN (hesb2)"  
/protein\_id="CA14937.1"  
/db\_xref="GI:3861037"  
/translation="MKNVISLDAAKQVLLIEKRAKPTFGIRVIGSGCAGQTY  
VEYADNKNQDFEYVEKEGVRLLIDPKTLMYTLGSEMDVETNFKISQFTFTPNKNKNC  
GCGKSFY"  
complement(4969. .5364)  
/gene="RP485"  
complement(4969. .5364)  
/gene="RP485"  
/codon\_start=1  
/transl\_table=1  
/product="NIFU PROTEIN (nifu)"  
/protein\_id="CA14938.1"  
/db\_xref="GI:3861038"  
/translation="MAYSKVIVDHYENPRNVSLDKKKKNVGTGLVGPACGDYMKQ  
IEVGDEITITAKRTFEGCSAISSSLVTFEMWIGKSVDEAKEIKNEIKELSLPPV  
KLHGSILAEADIKAAIADYKOKREKKDS"  
complement(5430. .6662)  
/gene="RP486"  
complement(5430. .6662)  
/gene="RP486"  
/codon\_start=1  
/transl\_table=1  
/product="NIFS PROTEIN HOMOLOG (spl1)"  
/protein\_id="CA14939.1"  
/db\_xref="GI:3861039"  
/translation="MNOCLKNLPLIYMDYSTPIPDPRVEMALPYFTTFCGNPHSR  
SHSFGMEANVENARSLVAVIGADSKKEIFTSGAVESNNLVYKGLAFYGNKKHII  
ITIVSEHKCVLNACRHLRBOBKITTYLPDKNGIITDETLKNAITDOTLLVSWAVNN  
EIVGIQPLKEIGKICREBNVFFHSDIAGQFKTPIINVECHIDLASISGKHTYPKGI  
GALYIRKPRVAVTPIPLNGGOERGMRGTLPTPLVGLGIASISIAINEMEKTOHNV  
YLDFRFLNNHISKISEVYLNGDKDQRYKGNLSPFAGVEGSIILAIKDLAVSSGSAC  
TSASLEPSYVLRISIGISEELAHTSIRFGIGRFTTEQEIADVAVNLVCSKIDKLRLSPL  
WEMMOEGVDLKKIRMTAH"  
complement(6740. .7870)  
/gene="RP487"  
complement(6740. .7870)  
/gene="RP487"  
/codon\_start=1  
/transl\_table=1  
/product="NIFS PROTEIN HOMOLOG (spl1)"  
/protein\_id="CA14940.1"  
/db\_xref="GI:3861040"  
/translation="MIYLDHNATTFIDPRVKEYIISLMDKELNPSAHTSGRFANKII  
ETASQIATSGITMSREYNTFTSSGTSNNLTMKNPYDGDIFISAIEHLSTYHNIN  
YAPNKITIRVAGKGLVDEHEDLDSGNSKRLVSMVANNENGVLNDIAAISIKIR  
KYNAKPSIDVOSGCKISINKEIGELDPATISGKIGGCHCCGALLINSNPOLPIII  
GGCGEKYSRSTENVLAIAGGLAEERRRDISNRTKIRIYDRLEROKLKEYSNVMI  
ISNNVARANTLITVNDTQAVRLIGFDLRNICVSSGACSSGKISKSHVLTNMGIR  
EEEARSSIRISLSTHTNTISDIKAFIKAFEEIYEYKSSCNHCI"  
complement(7931. .8125)  
/gene="RP488"  
complement(7931. .8125)  
/gene="RP488"  
/codon\_start=1  
/transl\_table=1





Db 23321 GTACAACCTCGTAAATCCGGTAGCATTTTACAA 23384

RESULT 37

RPMX04 237523 bp DNA linear BCT 11-NOV-1998

LOCUS Rickettsia prowazekii strain Madrid E, complete genome; segment 4/4.

ACCESSION AJ235273 AJ235269

VERSION AJ235273.1 GI:3861237

KEYWORDS complete genome.

SOURCE Rickettsia prowazekii.

ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.

REFERENCE 1 (bases 1 to 237523)

Anderesson, S.G., Zomrodipour, A., Anderesson, J.O., Sichteritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H. and Kurland, C.G. The genome sequence of Rickettsia prowazekii and the origin of mitochondria

TITLE Nature 396 (6707), 133-140 (1998)

JOURNAL 9823893

MEDLINE 9823893

PUBMED 9823893

REFERENCE 2 (bases 1 to 237523)

Anderesson, S.G.E. Direct Submission

AUTHORS Submitted (11-NOV-1998) S.G.E. Anderesson, S.G.E. Anderesson@molbio.uu.se, Dept. of Molecular Biology, University of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN

JOURNAL Location/Qualifiers

FEATURES

Source 1..237523

/organism="Rickettsia prowazekii"

/strain="Madrid E"

/db\_xref="taxon:782"

164..1123

/gene="Rp701"

164..1123

/gene="Rp701"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="CA15137.1"

/db\_xref="GI:3861238"

/translation="MAEPIIFTLFGKRDILDRINNOKALNKLHIIPLOVKNESHISDKERISEILEKSYKECKIPFIENIOISNNVKKPIFRAODLIINAONLKITTEDEKNALPONSLEKSYMOIIEKVDHVFETNEADONLSIADGIVPKDKTTTIDIRLTSVFNLIYNHETNHLGSLTIKKYDKLIKTKAKKODSKELIETPISVDEANTLITAKGITS EDOYIGLKEIHELKDPNNAENFKYVQISQPKDKDKOKKHIDSNFNTQKVI NDSNKDIOVEHITSYVODKERQSKPGFLTRIFENYKNIYVKO"

2074..3348

/gene="Rp702"

2074..3348

/gene="Rp702"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="CA15138.1"

/db\_xref="GI:3861239"

/translation="MPTLIIVIIIMISLSALFAETFAITASSPKIHKLTKAGNR AKTVLEVLKKKKEVIGILLIGNSLINTICITITATLFLISLGNNGTIVAGVAFIT I VRAVVPKATAVAKPEOLTLKMSSTIVFLKRPINIAIDYITKJPCFLFRINLPO I ISGTEEVKVI EHYHOGGVYKSERNNGIILDIRNMTVSEIMTHRSNIITANIDIPY ELIITKTLISGAHRIPLMDNRDNIIGILNKLKALAYENNDKQKVDVITLTPPM FTIDNALVVDQLHAERNNHAGVDEYGLTGLITLEDVIEIVGPIIDDEHRIINN EITQESNTERIIGTITIRIDINRELDMNLSDEPANTAGLIHKIARIIPNOGEVITK FNEKIIILKRIANKIDSVKITVLPMTETTISSE"

3341..5353

/gene="Rp703"

3341..5353

/gene="Rp703"

/codon\_start=1

/transl\_table=11

/product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCME (ccmf)"

/protein\_id="CA15139.1"

/db\_xref="GI:3861240"

/translation="MSKIGNFLITTCISGLITLITLIPMLDLVILKKHNNCLKRF PMIPCTISQVMTQATETFASSLSCTIFACTLVAVETISDEVONVPLHSSLPKLITKI SGWSHSEHGMILMFLCLOIVSCYIECLDKMLKRSIIILLSAIFLAVISRDN PFHMFSEVPEKGLNPMLODIALSHIPLLLYGVYAVPTIVCAILFIAVISRLD NMLPGYPTGKRIOTOLSDRITSIRLKFACGILEFTTIGISGMAVRELNGLMFE FDEPENSLEPMILSGIMLHHSITVTRKTNMOHMTIILSTIFLIVFSEFLVRSQGI NSIHAFAPFSSERNIYLIVITICIGLYIYIRLNNSCIERTYNTFNTRHSCNMT IKERAILIGNIFPFLSLILICAPLVPTIYALNDKPTINIEEFIKNEITIEVPLIC TIGLETTNNSIKKHIIIMLSLITVFLINKEFSVLSITITMSIFLMAHVAIYLV KTYVFRNSKASTMSIIGHFVALLAFSTTMSLSQSEMDFTGEVTSKTFNAPFVLT LONIKFQAGKNYROIAPFMLEHDKNNITILKPEENREYIIEOKLSQESDIHSLVDL YAVLSNDGKIHAKIYKPMSPFIMGVLITTAGFVIALIRNSV"

complement(12764..12195)

/gene="Rp704"

complement(12764..12195)

/gene="Rp704"

/codon\_start=1

/transl\_table=11

/product="CELL SURFACE ANTIGEN (scas)"

/protein\_id="CA15140.1"

/db\_xref="GI:3861241"

/translation="MAQKPNFLKRIISAGLVATSTATVAGFSGVAMGAMOYRNTN AAATTEGIGFDOAGANIPVAPNSVITTAANNPIFENPENGHLNLFETDADLAVT INDDTGLFTITNIAOAKPFNFVPAKGLINITIGOGITVGEANTTINAONALIKVHG AAINANDSLGISTEPAAPSVLEPNLINTPOEALITGANSKIVNGSGTNTLNTG FIVSDNTPAGCITINIDCGGLMNSTPDAANTLNOVGSGNTINRNGIDGCKLVLY SKNGATEFNVGTGLGNLKGIIELNAAVAGLISQGGAAVNICTDNGSAGAAPIE VSDVNGNAATISQOYAKMNVIOSANAGVTEHEIYDVLGGTTEKADSVKIIIE NSMFGSTENFNGLDTOIIVPDKILKFGIDVNNNGNTAGVIFPNAGVLSASTDPN IAVTNINAIAPGAGVLESGIHAEIRLNGSGSIRKALGVINGPNOVALMNNNA LAAGSTOLOGSATITIDGNGVNAALOHTTLANDSKILALIDAGNACVAGLAE OANGCTIKLTNTONNIVNFDLITTDKTVADVASSITNNQITTINGISCTYANRKT LAOLNIGSSKTTILNAGDVALNEVLTEKNSVOLNHNVTILTKTTINANGQIIIVADP LNTNTTLADGTNLGSAENPLSTHFAKTKANASLIANGVNLVANNITNTDANVS LHFRESGTSIVSGTGGQOQGHKLNMLLDNGTVKFLGDTENGGRKEGKSILQISN NYTTHVESASDNTGTLFENVDPITVTLKNOGAFEGVLKOVILISGNTVFNNGVGS IYHGIANSISPEFNASISGTSFLPISGTPLDVLTIKRGTGNGVLDNPNAPLVNVSIDS MINNGOIIQDKKNIILALSQSDNSTYNNANTLSGIRTRKNNQGYTTLSEGMARNNGT IYGLGLENGSPKLNQVFTTDDYNNLSIILANNVTINDVYTLTGGIAGDPEDKITLIG SVGNANVREVPDSTFSDPRSMIVATANKCTVYLDNALVSNITDTPVAVSRFGN DSGAGLOGNITVSONIDFGTYNLTILNSVYLLGGTVAINGEIDLNTNMLIFNGTSTW GDNSTSTLTNVSNGNIGOVYLAEDOVANATGGTTIKIDOMANNEFGTOVAYTLIO GGRFNGTGAAPVATGSMIEPKYELIRPSNDQVATRPNDVLANVVTAVVANGSALAN APCVSONISRSCLESTMAAYNNMLAKDPSDVAETPVGATITRDSAAVTYVANNIDPKT ODLNSNLGTLIRLSNAETSDVAGSATGAVSSGEDEVSTGWAKFPYINAEODKRG IAGYKAKTGVVGLDTLANSNDMLIGAAGITVTDIRHODYKGDGTDOJNGLSFSLYG SOOLVNFEPAGNAIFTLNKKRSQRPESGKRSKQIAAONDMNTEFGNLLISYG DYANMPVLTVPMAGLSLYKSNENYKEGTGYANKRINSKFEEDRYDLIVGAIVAST VNTIDIVIVEIHSFVNHKYNKLSNSQSLDQGTAPFISQDPRTAKTSYVIGLSANI KSDAKMEYIGCYRPNASKTYTAHQGLKRVN"

complement(12953..13216)

/gene="Rp705"

complement(12953..13216)

/gene="Rp705"

/note="Pseudogene that lacks ATG or GTG initiation codon. Homology to guanosine 3'-5'-bis(diphosphate) 3'-pyrophosphohydrolase."

complement(13563..14507)

/gene="Rp706"

complement(13563..14507)

/gene="Rp706"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="CA15141.1"

/db\_xref="GI:3861242"

/translation="MKIROTAKPVITIGISPELTDARELFEENHPLGIIILFRNRIRK NEKGEDQKALIKLADIKEVGLDNTIINIDGGGVYKRLIAPTPYDADAOTFTELO TCKYNSYTIKELREVGINDPAPVADLIHDGADPKYIISDRSFGKEPEIVVPLFLSAIA







Db	4290	TTGTCCTTGTACTATATGATTTGACCAACAATTTTATCATTTATTTGATGGAGCAAT	4349
QY	1221	ACAA---CTCCTCATGCTACTGTGGCTGGCGAGCATATACAGAAATTTCTTCGGAAA	1277
Db	4350	TTAAATGTCACCTCTGTCTGTGGCGCTTTCAATGGGGGATTTCAAAATTTTATTTAGCTTAA	4409
QY	1278	TCACAAATATACGCTCTCTTTGACTCAACTAAAGAANAATGGCTATATCCCTTTGACCAA	1337
Db	4410	GGTACTAATATTTATATATGGGATACTTCACAGAAATTTATATATCCCTCTTGATGAG	4469
QY	1338	GAGCAAAATGTCAAAGGTAAAGGCTGCTATGTAGTTGATTTGCCGCCGCTTGCGAAATCA	1397
Db	4470	GAGCTTAAACCAAGCAAGCAAGCTCGGTGACGTGATAGTACAAAGTCCGCAAAATCC	4529
QY	1398	GGAGGAGCTTTAATCCAA	1415
Db	4530	TCCTAGCGCTTTGTGCAA	4547
RESULT	39		
LOCUS	AC018848	132990 bp	DNA linear
DEFINITION	Arabidopsis thaliana chromosome 1 BAC F516 genomic sequence,		
ACCESSION	AC018848		
VERSION	AC018848.6		
KEYWORDS	HMG.		
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana.		
REFERENCE	Arabidopsis thaliana. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 132990)		
AUTHORS	Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J., Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Bainsread,M.E., Bowman,C.H., White,O., Nierman,W.C. and Fraser,C.M.		
TITLE	Arabidopsis thaliana chromosome 1 BAC F516 genomic sequence		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 132990)		
AUTHORS	Lin,X. and Kaul,S.		
JOURNAL	Direct Submission		
TITLE	Submitted (21-DEC-1999) The Institute for Genomic Research, 9712		
REFERENCE	Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org		
AUTHORS	3 (bases 1 to 132990)		
JOURNAL	Town,C.D. and Kaul,S.		
TITLE	Direct Submission		
COMMENT	Submitted (19-JAN-2001) The Institute for Genomic Research, 9712		
	Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org		
	On Jan 19, 2001 this sequence version replaced g1:12280825.		
	Address all correspondence to:at@tigr.org		

http://Ftp.genome.washington.edu/RM/RepeatMasker.html).

```
source
1..132990
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F516"
1..7662
/overlap with BAC clone F18b13
[AC009322:109243..116905]..
1..7662
/note="the annotation for genes identified in this region
can be found in the overlapping bac F18b13 sequence
109243-116905"
gene
/feature="complement(1535..1606)
/gene="60014.tRNA-Trp-1"
complement(1535..1606)
/gene="60014.tRNA-Trp-1"
/product="tRNA-Trp"
7875..10851
/gene="F516.2"
/note="contains Pfam profile: PF01535 Domain of unknown
function"
join(<8785..8813,9004..9517,9604..>10851)
/gene="F516.2"
join(8785..8813,9004..9517,9604..10851)
/gene="F516.2"
/codon_start=1
/product="hypothetical protein: 8785-10851"
/protein_id="AAG52431.1"
/db_xref="GI:12324975"
/transcript="MFALSKVLRRTTORLRGACSAVFSKDIOLGGERSPDSNIASTR
BEAVRFEIEISLSNRALSSAGTSKQDEEDLEDGSELEGSKSGGSPDSDEGK
LSADEEEEDLIDETDYSRKTVEKQSELEKTVASPGISALDKVNEENETR
VEIAAMOLQRRMYGPAIOMSEMLANKTEMPDYSRDLTVKTRGLEKGCAC
NOKTRSKGEVLRTLLANCVANGNPKSSLYNKKKIDGIFPLSGFTCDOMILLEKR
IDRRKIAVLVLLKEKENIKPSLTYKTLIDVKGATNIOSEMEDILETMDEGVELDQ
TQALRRYISGAGLKDKAKEYLKEBGESEANRARKDLSTVASIGRDEEVKRTWK
ICESPVEEESLAIIOAFGLNKKVQEALEFEKLVKDRASSSTYSVLLRIVVDKMS
LSGKQDLVKRMAESGCRLEATTMDALIKLVAVEGEVAKSLLDKASKOSHTKTKLMS
FMVIMDEXSKRGDVNTEKFIEMKREPGYGRSLRQALQMAQIYNKSPAYGNRDLK
ADNIFPKNSMAQLAQGDPEFKKTAISLDL"
complement(11216..13661)
/gene="F516.3"
/note="contains Pfam profile: PF00561 alpha/beta hydrolase
fold"
complement(join(<11216..11469,11550..12077,12183..12914
13002..13309,13397..>13661))
/gene="F516.3"
complement(join(11359..11469,11550..12077,12183..12914,
13002..13309,13397..13661))
/gene="F516.3"
/codon_start=1
/product="unknown protein: 13661-11359"
/protein_id="AAG52432.1"
/db_xref="GI:12324976"
/transcript="MGKGMFVEKARCLRPVFMVAMLYSLVSLPVLVAIGDVLV
PTFLISPTCLTCYGEKHLRSYAFKSLTDLIDPLYSIVRSFELTCYSLSDARALSHG
PYLGLTSCVSVVLSVSKACVFTANSQLDASSSPSKORLHLKSKMVPPLFSS
VVFALGHVWVARTSCRRARRLHNVRDPAVLSSCKSVFSGYQVRSPIPLVGNKSK
VDGEARLTHFSSNDGELFARLALDLSEFTVRLGVYKVICTPASPRHSISSTV
EANSLLNRPDEAMVGRKLDRIINISMTVRNKLNNHNRYSLSLENNSSSLHDLDP
TSPLRFKIOECSRGDGINVSNFGATDEDDGNGCGFVLLVHGGGCVFSNRHNS
SLANDGCVYATDRPFGCGLTARPKKIDDEERKPNPTLDNVDYMLLAFCEMKRAS
VVLGHDGGSGLAKAQRLETKDPLKVGUVLLNSLTREVVPAPFLLHTSLGK
KHLVRLRTIEIAQVNNRWAYDPKATTTVLRKYARPLVECDMALEIRLSSEM
VLTPONALSLTAEAVENLPVLVAGADALVPLRSQVMASKLENSRLVAISCGHLPH
ECCPAILLAACPSPTLSRFSFD"
gene
16105..17094
/gene="F516.4"
<16105..>17094
/gene="F516.4"
```

Query Match	Best Local Similarity	8.5%;	Score 139.8;	DB 8;	Length 132990;
Matches 261;	Conservative	0;	Mismatches 202;	Indels 0;	Gaps 0;
QY 121	TAAGAAGTTCGCCAATGTTCTTAATGTCCTTCGTATTTACATTTTACATACGCTGT	180			
Db 19811	TGAAAAGATTTATTCCTTTAGGATGATGTTCTTTGATTTCTTTCAATTAACAATTC	19870			
QY 181	TACGGATACCAAAACACACTCTATTGGGAGCTCCGTGGTTCTGGTCAGAGCAATAC	240			
Db 19871	TGAGGGAATCAAAAGATGCTTGGGGGAGCGCGCAAAAGAAATTCCTGAGATTATAC	19930			
QY 241	CTTTCATCAAGTTTGGCTGTTGTGCCCTGGCTGCTATATCTTTATGCTTTATATGCA	300			
Db 19931	CTTTCTTGAGACTTGGGTGAATCTTCTCATGGCCATTGGGTTTATGCTCCTACACTA	19990			
QY 301	AGCTAAGTATATTTTAAGTACAGAGGCTTATTTTATGCACTGGGAAAGCCCTTTTAA	360			
Db 19991	AACCTCCAAATGTTCTCTCCAGAAGAGCGCTGTTTACACATGTTATGTCCCTTCATCA	20050			
QY 361	TTTTCTTGGCCTGTTCGCCAGCTGAATTTATCCGCTACGGATGTTTACATCTACAG	420			
Db 20051	TCTACTTTGGGGCCCTTGGTTTCGCAATACCCCTCAGCACTATATTCACCCGGAG	20110			
QY 421	AATTTCTGACCGTTTACAGGCCATCTACACCCTCCAGATTTGGTAGACATCGTTGACAT	480			
Db 20111	CTCTGCGAGATTAAGCTCCTTACACACCTCGGCCCAAGATTCAATGGCTCTATTGCAATAT	20170			
QY 481	TAGAAGCTGCAGCAATTCCTGCACTTTATATGACTTGCTGAACATATGGGAAAGCGTCAAC	540			
Db 20171	TGCGGATTTGGAGTTTCTGTTTGTATTTATGTTATGCTTGAGCCTTGGGGTATGGTGG	20230			
QY 541	TATCTCTAATGTTCTGGGGATTTGCTAATGAATTAACAAAAT	583			
Db 20231	TCTCAGTTCTCTCTGGGGCTTGTCTAATCAGGTGACTTAACACT	20273			
RESULT 40					
AC126791					
LOCUS					
DEFINITION	AC126791	217975 bp	DNA	linear	HTG 23-AUG-2002
ACCESSION	AC126791				
VERSION	AC126791.11	GI:22450616			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	barrel medic.				
ORGANISM	Medicago truncatula				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;				
	Medicago.				
REFERENCE	1 (bases 1 to 217975)				
AUTHORS	Shuall,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,				
	Cook,D., Kim,D. and Roe,B.A.				
TITLE	Medicago truncatula BAC Clone mtl1-7k2				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 217975)				
AUTHORS	Shuall,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,				
	Cook,D., Kim,D. and Roe,B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-JUN-2002) Department Of Chemistry And Biochemistry,				
	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
	OK 73019, USA				
REFERENCE	3 (bases 1 to 217975)				
AUTHORS	Shuall,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,				
	Cook,D., Kim,D. and Roe,B.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-AUG-2002) Department Of Chemistry And Biochemistry,				
	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
	OK 73019, USA				



## COMMENT

On Aug 23, 2002 this sequence version replaced g1:22417350.  
----- Genome Center  
Center: Department of Chemistry And Biochemistry  
The University of Oklahoma  
Center code:UOKNOR

NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2033: contig of 2033 bp in length  
\* 2034 2133: gap of unknown length  
\* 2134 4316: contig of 2183 bp in length  
\* 4317 4416: gap of unknown length  
\* 4417 7180: contig of 2764 bp in length  
\* 7181 7280: gap of unknown length  
\* 7281 9824: contig of 2544 bp in length  
\* 9825 9924: gap of unknown length  
\* 12804 12903: contig of 2879 bp in length  
\* 12904 15185: contig of 2282 bp in length  
\* 15186 15285: gap of unknown length  
\* 15286 17545: contig of 2260 bp in length  
\* 17546 17645: gap of unknown length  
\* 17646 20257: contig of 2612 bp in length  
\* 20258 20357: gap of unknown length  
\* 20358 23699: contig of 3342 bp in length  
\* 23700 23799: gap of unknown length  
\* 23800 26276: contig of 2477 bp in length  
\* 26277 26376: gap of unknown length  
\* 26377 29259: contig of 2883 bp in length  
\* 29260 29359: gap of unknown length  
\* 29360 32359: contig of 3000 bp in length  
\* 32360 32459: gap of unknown length  
\* 32460 36613: contig of 4154 bp in length  
\* 36614 39318: contig of 2605 bp in length  
\* 39319 39418: gap of unknown length  
\* 39419 41836: contig of 2418 bp in length  
\* 41837 41936: gap of unknown length  
\* 41937 45761: contig of 3825 bp in length  
\* 45762 45861: gap of unknown length  
\* 45862 51599: contig of 5738 bp in length  
\* 51600 51699: gap of unknown length  
\* 51700 58201: contig of 6502 bp in length  
\* 58202 58301: gap of unknown length  
\* 58302 64490: contig of 6189 bp in length  
\* 64491 64590: gap of unknown length  
\* 64591 70460: contig of 5870 bp in length  
\* 70461 70560: gap of unknown length  
\* 70561 75838: contig of 5278 bp in length  
\* 75839 75938: gap of unknown length  
\* 75939 82428: contig of 6490 bp in length  
\* 82429 82528: gap of unknown length  
\* 82529 87785: contig of 5257 bp in length  
\* 87786 87885: gap of unknown length  
\* 87886 95365: contig of 7480 bp in length  
\* 95366 95465: gap of unknown length  
\* 95466 102528: contig of 7063 bp in length  
\* 102529 124627: gap of unknown length  
\* 124628 124527: contig of 21799 bp in length  
\* 124528 169060: gap of unknown length  
\* 169061 217975: gap of unknown length  
\* 169161 217975: contig of 48815 bp in length.

## FEATURES

location/Qualifiers  
1..217975  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"

/clone="mchl-7k2"  
BASE COUNT 71744 a 35163 c 35694 g 72650 t 2724 others  
ORIGIN

Query Match 8.4% Score 138; DB 2; Length 217975;  
Best Local Similarity 59.2%; Pred. No. 1.5e-22;  
Matches 271; Conservative 0; Mismatches 185; Indels 2; Gaps 2;

OY 120 CTAAGAAGATGTCGCCAAGTTCCTATGTCCTGCTATACATTAACATATACGCTG 179  
Db 50970 CTCAGAAGATTAATACCACTGGAGATGATGTTTGTATGTTGTCATATACCAATT 51029  
OY 180 TTACGCATACAAAGACACTTATTGTGGAGCTCCTGCTGTCGAGAGCAATA 239  
Db 51030 CTAGAGATACAAAGATGTTCTGTTGTGACCTGAGAGAGATGCTGACATATATA 51089  
OY 240 CCTTCATCAAGTTTGGCTTGTGTCCTGCTGCTATTAATCTTAATGCA 299  
Db 51090 CCTTTTGAAGACATGGGTGAATCTTCATGCTATGCTTATGTTGTGACACA 51149  
OY 300 AAGCTA-AGTAATATTAATTAAGACAGCCTTATTTATGCACTGGAGCCCTTTT 358  
Db 51150 AAGTATTCCTATATGTTTGTCAAAACAGGCTTTTATCTGTATGTCCTTTAT 51209  
OY 359 AATTTCTTCCTGCTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418  
Db 51210 TGCCTTCTGGGCTTTGGGCTTTGTTGTATGCTGCTGCTGCTGCTGCTGCTGCT 51269  
OY 419 AGAATTCGTGACCGTTTACAGGCCATCTACCTCAGAGATGCTAGGACTGTGCAT 478  
Db 51270 GGCATTTTG-TGATTAACCTTCTTATATATCTGACCTGCTGCTGCTGCTGCTAT 51328  
OY 479 CTTAGAAACGACATTTGCTGCTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 538  
Db 51329 TATGAGATTTGAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 51388  
OY 539 GCTATCTCTAATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576  
Db 51389 GATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 51426

Search completed: January 21, 2003, 19:36:30  
Job time : 5227 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 18:03:28 : Search time 59 Seconds  
(Without alignments)  
8508.981 Million cell updates/sec

Title: US-09-869-433-1

Perfect score: 1637  
Sequence: 1 gaataaataacatacagaa.....gttcacccactcgtcttt 1637

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/lna/PCrus.COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.4	2.9	7218	1 US-08-232-463-14	Sequence 14, Appl
2	38.8	2.4	2718	4 US-09-074-658-69	Sequence 69, Appl
3	38.4	2.3	1269	4 US-09-134-001C-1211	Sequence 1211, Ap
4	38	2.3	1017	4 US-09-134-001C-2318	Sequence 2318, Ap
5	35.8	2.2	17949	4 US-09-087-465-3	Sequence 3, Appl1
6	35.4	2.2	1690	1 US-08-276-452A-24	Sequence 24, Appl
7	35.4	2.2	1690	2 US-08-798-744-24	Sequence 24, Appl
8	34.6	2.1	1647	1 US-08-198-446B-9	Sequence 9, Appl1
9	34.6	2.1	1647	2 US-08-870-693-9	Sequence 9, Appl1
10	34.4	2.1	729	4 US-09-134-001C-1161	Sequence 1161, Ap
11	33.8	2.1	658	4 US-08-998-416-595	Sequence 595, App
12	33.6	2.1	3946	4 US-09-453-702B-103	Sequence 103, App
13	33.4	2.0	405	4 US-09-134-001C-980	Sequence 980, App
14	33.4	2.0	585	4 US-09-222-939-30	Sequence 30, Appl
15	33.4	2.0	2590	3 US-08-714-918-78	Sequence 78, Appl
16	33.4	2.0	2590	4 US-09-265-315-78	Sequence 78, Appl
17	33.4	2.0	2590	4 US-09-265-315-78	Sequence 78, Appl
18	33.4	2.0	2590	4 US-09-265-315-78	Sequence 78, Appl
19	33.4	2.0	2706	4 US-09-463-238-2	Sequence 2, Appl1
20	33.4	2.0	2806	4 US-09-463-238-11	Sequence 11, Appl
21	33.4	2.0	3483	4 US-09-130-491-3	Sequence 3, Appl1
22	33.4	2.0	17410	1 US-07-841-646-3	Sequence 3, Appl1
23	33.4	2.0	17410	1 US-08-147-023-3	Sequence 3, Appl1
24	33.4	2.0	17410	1 US-08-447-570-3	Sequence 3, Appl1
25	33.4	2.0	17410	2 US-08-449-700-3	Sequence 3, Appl1
26	33.4	2.0	17410	2 US-08-449-699A-3	Sequence 3, Appl1
27	33.4	2.0	17415	3 US-08-486-343A-1	Sequence 1, Appl1

c	28	33.4	2.0	17415	5 PCT-US95-07349-1	Sequence 1, Appl1
	29	33.2	2.0	1691	3 US-09-013-881-12	Sequence 12, Appl
	30	33.2	2.0	3989	1 US-08-327-494A-1	Sequence 1, Appl1
	31	33.2	2.0	3989	1 US-08-327-494A-3	Sequence 3, Appl1
	32	33.2	2.0	3989	5 PCT-US95-13659-1	Sequence 1, Appl1
	33	33.2	2.0	3989	5 PCT-US95-13659-3	Sequence 3, Appl1
	34	33.2	2.0	5163	3 US-08-700-651-1	Sequence 1, Appl1
	35	33.2	2.0	5163	3 US-08-928-361B-4	Sequence 4, Appl1
	36	33.2	2.0	5318	3 US-08-700-651-2	Sequence 2, Appl1
	37	33.2	2.0	5318	3 US-08-928-361B-3	Sequence 3, Appl1
	38	33	2.0	200	4 US-09-020-956-140	Sequence 140, App
	39	33	2.0	200	4 US-09-030-607-140	Sequence 140, App
	40	33	2.0	200	4 US-09-605-785-140	Sequence 140, App
	41	33	2.0	200	4 US-09-429-313-140	Sequence 140, App
	42	33	2.0	200	4 US-09-352-616A-140	Sequence 140, App
	43	33	2.0	200	4 US-09-232-149A-140	Sequence 140, App
	44	33	2.0	2220	4 US-08-934-386-2	Sequence 2, Appl1
	45	33	2.0	4619	2 US-08-874-186-38	Sequence 38, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHRIFFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOULPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14  
Query Match 2.9%, Score 47.4, DB 1, Length 7218;

Best Local Similarity 5.3%; Pred. No. 0.0028;  
Matches 21; Conservative 210; Mismatches 166; Indels 0; Gaps 0;

```
OY 130 TTTCGCCAATGTCCTTAATGTTCTCTGATATACATTTAACTATAGCGGTGATACGCGATA 189
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1080 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1139
OY 130 CAAAGACACTTATTTGTGGAGCTCCTGTTGTTGTCGACAGGCAATACCTTTTCATCA 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1140 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1199
OY 250 AGTTTGCTGTGTGCCCTGTGCTATATCTATATGCTATATTAATGCAAGCAATCA 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1200 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1259
OY 310 ATATTTAAGTAGCAGCGCTTATTTATTAAGCAGCGCCCTTTTAAATTTCTTTG 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1260 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1319
OY 370 CCTGTTCGCCAGCTGAATTTATCCGTAAGCGGATGTTTACATCCATACAGATTTGCTG 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1320 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1379
OY 430 ACCGTTACAGGCAATCCCTACCTCCAGATGCTAGACTGCTGCTTCCATCTTAAGAACT 489
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1380 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1439
OY 490 GGACATTTGCTGCAATTTATGCTGCTGCTGCAATATG 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1440 AATCTCTCTATCTCTTTAACTACTGATGATAGG 1476
```

RESULT 2  
US-09-074-658-69/c  
; Sequence 69, Application US/09074658  
; Patent No. 6184371  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M  
; APPLICANT: Run-Pan Du  
; APPLICANT: Quijun Wang  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
; NUMBER OF SEQUENCES: 78  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/074,658  
; FILING DATE: 08-MAY-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-795  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 69:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2718 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: Linear

US-09-074-658-69

Query Match 2.4%; Score 38.8; DB 4; Length 2718;  
Best Local Similarity 57.4%; Pred. No. 0.44;  
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
OY 1079 TGTGTGCTGTAACGTCAATGTAATTTGGATGCTTAACGTGAGCCCTAGTCACTCCTGT 1138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1231 TATGTGCTGTATCATCATCAAGATTTGGCTGTGCCAGTAAGCGGTTGCCATTAATCT 1172
OY 1139 CATGCTTCTCTACAGGTATGCTTTCTGCTGCTTCTTATCTTTAAGAACCAAGCTTC 1198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1171 TGGGCTCATCTCATGACATCGGCTTCACTTGTGCTTTTGTGCTATCAACATGACTTC 1112
OY 1199 TG 1200
      : :
Db 1111 TG 1110
```

RESULT 3

US-09-134-001C-1211  
; Sequence 1211, Application US/09134001C  
; Patent No. 6380370

GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: CTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1211  
; LENGTH: 1269  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1211

Query Match 2.3%; Score 38.4; DB 4; Length 1269;  
Best Local Similarity 47.5%; Pred. No. 0.41;  
Matches 114; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```
OY 1019 TGAATTCATGGGAACTTCCTCTCGACGCGGTAGTATCCGTAATTCATGATATT 1078
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 885 TAAATTAACAAAATTTATACCTACATTTCTTTACTATTTTACAAATCACTACTATT 944
OY 1079 TGTGTGCTGTAACGTCAATTCGTAATTTGGATGCTTAACGTGAGCCCTAGTCACTCCTGT 1138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 945 TGTACTGTATTAACAATTAATCAATATAGCATTAATCATTTGCTTTATTCGGTTTGG 1004
OY 1139 CATGCTTCTCTACAGGTATGCTTTCTGCTGCTTCTTATCTTTAAGAACCAAGCTTC 1198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1005 TACATTTGGTAACACCGCTTACTTAATAGCAAAATATATCTTAAGTCAAAAGAACACC 1064
OY 1199 TGGGCTGTGCGCTATGTTGCTGATCACTGCTCATGCTAGCTGTGCTGTGAGGATAT 1258
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1065 ACTTCTTGCAAGTACGTTAGCAGCTTCTATTTTCAATGTTGCTAATTTCTTGGTGAAT 1124
```

RESULT 4

US-09-134-001C-2318  
; Sequence 2318, Application US/09134001C  
; Patent No. 6380370

GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: CTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; FILING DATE: 1998-08-13





RESULT 9  
 US-08-870-693-9/c  
 Sequence 9, Application US/08870693  
 Patent No. 5866538  
 GENERAL INFORMATION:  
 APPLICANT: Hartwell, Leland H.  
 APPLICANT: Weinert, Ted A.  
 APPLICANT: Plon, Sharon E.  
 APPLICANT: Groudine, Mark T.  
 TITLE OF INVENTION: Cell Cycle Checkpoint Genes  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
 STREET: 1420 Fifth Ave., Suite 2800  
 CITY: Seattle  
 STATE: WA  
 COUNTRY: USA  
 ZIP: 98101-2347  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS

```

RESULT 10
US-09-134-001C-1161/c
: Sequence 1161, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GFC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: PRIORITY FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIORITY FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIORITY FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1161
: LENGTH: 729
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1161

```

Query Match	2.1%;	Score 34.4;	DB 4;	Length 729;
Best Local Similarity	47.3%;			
Matches 104; Conservative	0;	Mismatches 116;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 11
US-08-998-416-595/C
Sequence 595, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Robischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII
TITLE OF INVENTION: AND USUS THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwalls Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8567
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 595:
SEQUENCE CHARACTERISTICS:
LENGTH: 658 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1408RP
US-08-998-416-595

```

Query Match	2.1%	Score 33.8;	DB 4;	Length 658;
Best Local Similarity	49.7%	Pred. No. 5.8;		
Matches 86;	Conservative 0;	Mismatches 87;	Indels 0;	Gaps 0;

Oy	2	AATTAATAAACTATACAGATTGAAAATTTAAAGTATTTTCAGGGGTAAATATGCACAAAAAC	61
Dd	611	ATTAATATTAATAAATAATTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	552D
Oy	62	CGAAGAAAAACCTTTTGGAATATTCGCGCTCTTCTGTGGCCGATACATPCTCAGCACT	12LD
Dd	551	ATTAATATTAATAATAGAAATAGAAATAGCATATATATATATATGTTCCACAATAATCAATCTATAT	492D
Oy	122	AAAGCAAGTCTTGCCAATGTCTCTAATGTTCTTCTGTATTACATTTAACTATA	174
Dd	491	AATGATTAATAAGTCACCATAATATATACATCTACTATGATATATATATATATATA	439

```

1 RESULT 12
2 US-09-453-702B-103/c
3 : Sequence 103, Application US/09453702B
4 : Patent No. 6365723
5 :
6 : GENERAL INFORMATION:
7 : APPLICANT: Blattner, Frederick R.
8 : Burland, Valerie
9 : Perna, Nicole T.
10 : Plunkett, Guy
11 : Welch, Rod
12 :
13 TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
14 :
15 NUMBER OF SEQUENCES: 265
16 :
17 CORRESPONDENCE ADDRESS:
18 : ADDRESSEE: Quarles & Brady
19 : STREET: 1 South Plinckney Street
20 : CITY: Madison
21 : STATE: WI
22 :
23 COUNTRY: US
24 :
25 ZIP: 53701-2113
26 :
27 COMPUTER READABLE FORM:
28 : MEDIUM TYPE: Diskette, 3.50 Inch, 1.44mb storage
29 :
30 COMPUTER: IBM PC compatible
31 : OPERATING SYSTEM: PC-DOS/MS-DOS
32 : SOFTWARE: Word Perfect 8.0
33 :
34 CURRENT APPLICATION DATA:
35 : APPLICATION NUMBER: US/09/453.702B
36 : FILING DATE: 03-Dec-1999
37 : CLASSIFICATION: <Unknown>
38 :
39 PRIOR APPLICATION DATA:
40 : APPLICATION NUMBER: 60/110,955
41 : FILING DATE: 04-DEC-1998
42 :
43 ATTORNEY/AGENT INFORMATION:
44 : NAME: Seay, Nicholas J.
45 : REGISTRATION NUMBER: 27386
46 : REFERENCE/DOCKET NUMBER: 960296.95017
47 :
48 TELECOMMUNICATION INFORMATION:
49 : TELEPHONE: (608) 251-5000
50 : TELEFAX: (608) 251-9166
51 :
52 INFORMATION FOR SEQ. ID NO: 103:
53 : SEQUENCE CHARACTERISTICS:
54 : LENGTH: 3946
55 : TYPE: nucleic acid
56 : STRANDEDNESS: double
57 : TOPOLOGY: linear
58 :
59 MOLECULE TYPE: DNA (genomic)
60 :
61 SEQUENCE DESCRIPTION: SEQ ID NO: 103:
62 :
63 US-09-453-702B-103

```

Query Match	2.1%:	Score 33.6:	DB 4:	Length 3946:
Best Local Similarity	52.1%:	Pred. No. 14:		
Matches 75:	Conservative	0:	Mismatches 69:	Indels 0:
			Gaps	0:
QY 171	TATACGGCTTACCGGATACAAACACACTCTTATGTGGAGCGCTCGTGTGTGCA	230		
Db 1604	TATTCGGCTTTTAAGCCATACAGAACTACCGCATTCATTAGGTGGCATTTCTTGAA	1545		
QY 231	GAGGAATACCTTTCATCATCAAGTTTGGCTGTGTGCCCTGCGATATATCTTATGCTT	290		
Db 1544	AATCAATTGAATGAGGAGCAATTCATTTCTTGATGTATTCAGTACATTAATATGGGCGAGTT	1485		



```
US-09-265-315-78
; Sequence 78, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ying J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,315
; FILING DATE: March 9, 1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/714,918
; FILING DATE: September 13, 1996
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 240/247
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEO ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2590 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-265-315-78

Query Match      2.0%; Score 33.4; DB 4; Length 2590;
Best Local Similarity 52.6%; Pred. No. 13:
Matches 70; Conservative 1; Mismatches 62; Indels 0; Gaps 0;
```

```
US-09-265-315-78
; Sequence 78, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ying J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,315
; FILING DATE: March 9, 1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/714,918
; FILING DATE: September 13, 1996
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 240/247
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEO ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2590 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-265-315-78

Query Match      2.0%; Score 33.4; DB 4; Length 2590;
Best Local Similarity 52.6%; Pred. No. 13:
Matches 70; Conservative 1; Mismatches 62; Indels 0; Gaps 0;
```

RESULT 17

RESULT 18



US-09-266-417-78  
: Sequence 78, Application US/09266417  
: Patent No. 6228588  
: GENERAL INFORMATION:  
: APPLICANT: Benton, Bret  
: APPLICANT: Lee, Vling J.  
: APPLICANT: Malouin, Francois  
: APPLICANT: Martin, Patrick K.  
: APPLICANT: Schmid, Molly B.  
: APPLICANT: Sun, Dongxu  
: TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS  
: TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS  
: TITLE OF INVENTION: TARGET GENES  
: NUMBER OF SEQUENCES: 111  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Lyon & Lyon  
: STREET: 633 West Fifth Street  
: STREET: Suite 4700  
: CITY: Los Angeles  
: STATE: California  
: COUNTRY: U.S.A.  
: ZIP: 90071-2066  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
: MEDIUM TYPE: storage  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: IBM P.C. DOS 5.0  
: SOFTWARE: Word Perfect 5.1  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/266,417  
: FILING DATE: March 9, 1999  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 08/714,918  
: FILING DATE: September 13, 1996  
: APPLICATION NUMBER: 60/009,102  
: FILING DATE: December 22, 1995  
: APPLICATION NUMBER: 60/003,798  
: FILING DATE: September 15, 1995  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Warburg, Richard J.  
: REGISTRATION NUMBER: 32,327  
: REFERENCE/DOCKET NUMBER: 240/248  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (213) 488-1600  
: TELEFAX: (213) 955-0440  
: TELEX: 67-3510  
: INFORMATION FOR SEQ ID NO: 78:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 2590 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: US-09-266-417-78

Query Match 2.0%; Score 33.4; DB 4; Length 2590;  
Best Local Similarity 52.6%; Pred. No. 13;  
Matches 70; Conservative 1; Mismatches 62; Indels 0; Gaps 0;

QY 835 AAAAGGGGAAAAAGAGCTTAACCTAAATGATGATAAGAGTACCTTCCTATCTTG 894  
DB 1731 AAAATTGAAAGRAGATTTGAATTAGATATATAAAATAGGCAATTATTCGTTA 1790  
QY 895 ATAGATCTCTTATATCTTTTATTAATCTCTTGTTATTCGCTATGCGTATTCATTA 954  
DB 1791 TAAGTATTTTACATAGATTTTCAAGATATGTTGCTTTCATCTCATATTTGCTAAT 1850  
QY 955 ACTTAATGCAAGT 967  
DB 1851 TGTTAAGCTATGT 1863

RESULT 19

US-09-463-238-2/c  
: Sequence 2, Application US/09463238  
: Patent No. 6469230  
: GENERAL INFORMATION:  
: APPLICANT: Edwards, Elizabeth A  
: APPLICANT: Smith, Alison M  
: APPLICANT: Bustos Guillen, Regla  
: APPLICANT: Martin, Catherine R  
: APPLICANT: Plant Bioscience Limited  
: TITLE OF INVENTION: Starch Debranching Enzymes  
: FILE REFERENCE: 97.118  
: CURRENT APPLICATION NUMBER: US/09/463,238  
: CURRENT FILING DATE: 2000-01-21  
: PRIOR APPLICATION NUMBER: PCT/GB98/02280  
: PRIOR FILING DATE: 1998-07-30  
: PRIOR APPLICATION NUMBER: GB 9716185.5  
: PRIOR FILING DATE: 1997-07-31  
: NUMBER OF SEQ ID NOS: 30  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 2  
: LENGTH: 2706  
: TYPE: DNA  
: ORGANISM: Solanum tuberosum  
: FEATURE:  
: NAME/KEY: misc.feature  
: LOCATION: (2641)  
: OTHER INFORMATION: n = a or g or c or t  
: US-09-463-238-2

Query Match 2.0%; Score 33.4; DB 4; Length 2706;  
Best Local Similarity 53.0%; Pred. No. 13;  
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1460 TTATCTGCAGTATCTCTTTTCATCATGCTAATTTGGTTGCTTCGCACTAAGTT 1519  
DB 2706 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 2647  
QY 1520 AAACAACATATCTTTAGGCGCATCTGCTTAACAACAAGAGGCTCAAGAAGATTG 1579  
DB 2646 CACTATATTTTATTTTATATATAATGCAATTAATAACAAGGTATCTTATAAATCATG 2587  
QY 1580 AGCTCTGCTTC 1591  
DB 2586 AGCTTCTTCCTC 2575

RESULT 20  
US-09-463-238-11/c  
: Sequence 11, Application US/09463238  
: Patent No. 6469230  
: GENERAL INFORMATION:  
: APPLICANT: Edwards, Elizabeth A  
: APPLICANT: Smith, Alison M  
: APPLICANT: Bustos Guillen, Regla  
: APPLICANT: Martin, Catherine R  
: APPLICANT: Plant Bioscience Limited  
: TITLE OF INVENTION: Starch Debranching Enzymes  
: FILE REFERENCE: 97.118  
: CURRENT APPLICATION NUMBER: US/09/463,238  
: CURRENT FILING DATE: 2000-01-21  
: PRIOR APPLICATION NUMBER: PCT/GB98/02280  
: PRIOR FILING DATE: 1998-07-30  
: PRIOR APPLICATION NUMBER: GB 9716185.5  
: PRIOR FILING DATE: 1997-07-31  
: NUMBER OF SEQ ID NOS: 30  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 11  
: LENGTH: 2806  
: TYPE: DNA  
: ORGANISM: Solanum tuberosum  
: FEATURE:  
: NAME/KEY: misc.feature  
: LOCATION: (822, 826, 2707 and 2797)



TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3192..3730  
OTHER INFORMATION: /label= EXON-1  
OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10413..10414  
OTHER INFORMATION: /label= GAP-1  
OTHER INFORMATION: /note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10696..10891  
OTHER INFORMATION: /label= EXON-2  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10960..10961  
OTHER INFORMATION: /label= GAP-2  
OTHER INFORMATION: /note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11059..11211  
OTHER INFORMATION: /label= EXON-3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11351..11352  
OTHER INFORMATION: /label= GAP-3  
OTHER INFORMATION: /note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11420..11617  
OTHER INFORMATION: /label= EXON-4  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11721..11722  
OTHER INFORMATION: /label= GAP-4  
OTHER INFORMATION: /note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 13354..13436  
OTHER INFORMATION: /label= EXON-5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 15044..15160  
OTHER INFORMATION: /label= EXON-6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 17245..17410  
OTHER INFORMATION: /label= EXON-7  
US-07-841-646-3

Query Match 2.0%; Score 33.4; DB 1; length 17410;  
Best Local Similarity 62.7%; Pred. No. 29;  
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 3 AATAAAGACATTCAGATAGAAATATAAGATTTTCGAGGCGTAAATATGCAAAAACC 62  
DB 4401 AAAAAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATTAAC 4342

OY 63 GAAGAAAAACCTTTTGAAATT 85  
DB 4341 AAGAAAGAAAGAAAGAAAGAAATT 4319

RESULT 23  
US-08-147-023-3/c  
Sequence 3, Application US/08147023  
Patent No. 5468845  
GENERAL INFORMATION:  
APPLICANT: OPPERMANN, HERMANN  
APPLICANT: OZKAYNAK, ENGİN  
APPLICANT: KUBERASAMPATH, THANGAVEL  
APPLICANT: RUEGER, DAVID C.  
APPLICANT: PANG, ROY H.L.  
TITLE OF INVENTION: OSTEOGENIC DEVICES  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: 53 STATE STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: U.S.A.  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/147,023  
FILING DATE: 21-FEB-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 810,560  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 827,052  
FILING DATE: 28-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 660,162  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,988  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,849  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 616,374  
FILING DATE: 21-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 600,024  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 599,543  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 579,865  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 569,920  
FILING DATE: 20-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 483,913  
FILING DATE: 22-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 422,613  
FILING DATE: 17-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 315,342  
FILING DATE: 23-FEB-1989  
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 232,630
FILING DATE: 15-AUG-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 179,460
FILING DATE: 08-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R.
REGISTRATION NUMBER: 27,829
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 17410 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: 3192..3730
OTHER INFORMATION: /label= EXON-1
OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10413..10414
OTHER INFORMATION: /label= GAP-1
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN
THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 10696..10891
OTHER INFORMATION: /label= EXON-2
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10960..10961
OTHER INFORMATION: /label= GAP-2
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN
THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11059..11211
OTHER INFORMATION: /label= EXON-3
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11351..11352
OTHER INFORMATION: /label= GAP-3
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN
THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 11420..11617
OTHER INFORMATION: /label= EXON-4
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11721..11722
OTHER INFORMATION: /label= GAP-4
OTHER INFORMATION: /note= "APPROXIMATELY _____ BASES ARE ESTIMATED TO
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN
THIS SEQUENCE."
FEATURE:
NAME/KEY: exon
LOCATION: 13354..13436
OTHER INFORMATION: /label= EXON-5
FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160

```

```

: OTHER INFORMATION: /Label= EXON-6
FEATURE:
: NAME/KEY: exon
: LOCATION: 17245..17410
: OTHER INFORMATION: /label= EXON-7
US-08-147-023-3

Query Match      2.0% Score 33.4; DB 1; Length 17410;
Best Local Similarity 62.7%; Pred. No. 29;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY      3  AATAAAACATGTCAGATTAGAAAATTAAGAATTTTCAGAGGGTAAATATGTACAANAACC 62
        ||| |||| | ||| ||||| |||| | || ||| ||| |||
Db      4401 AAAAAAAAAAAAAAGAAAGAAAAGAAAAGAAAAAAGAAAAAAGAAAAAATTAAC 4342
              ||||| | |||||

OY      63  GAAGAAAAACCTTTTGAAAAATT 85
              ||||| | |||||

Db      4341 AAAGAAAGAAAACAAGAAAAATT 4319

RESULT 24
US-08-447-570-3/c
: Sequence 3, Application US/08447570
: Patent No. 5714589
GENERAL INFORMATION:
APPLICANT: OPPERMANN, HERMANN
APPLICANT: OKAYNAK, ENGIN
APPLICANT: KUBERASAMPATH, THANGAVEL
APPLICANT: RUEGER, DAVID C.
APPLICANT: PANG, ROY H.L.
TITLE OF INVENTION: OSTEOGENIC DEVICES
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HORWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,570
FILING DATE: 21-FEB-1992
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 810,560
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 827,052
FILING DATE: 28-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 660,162
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 621,988
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 621,849
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 616,374
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 600,024
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 599,543
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 579, 865  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 569, 920  
FILING DATE: 20-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 483, 913  
FILING DATE: 22-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 422, 613  
FILING DATE: 17-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 315, 342  
FILING DATE: 23-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 232, 630  
FILING DATE: 15-AUG-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 179, 460  
FILING DATE: 08-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER, EDMUND R.  
REGISTRATION NUMBER: 27, 829  
REFERENCE/DOCKET NUMBER: CRP-001CE6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
FAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17410 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3192..3730  
OTHER INFORMATION: /label= EXON-1  
OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10413..10414  
OTHER INFORMATION: /label= GAP-1  
OTHER INFORMATION: /note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10696..10891  
OTHER INFORMATION: /label= EXON-2  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10960..10961  
OTHER INFORMATION: /label= GAP-2  
OTHER INFORMATION: /note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11059..11211  
OTHER INFORMATION: /label= EXON-3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11351..11352  
OTHER INFORMATION: /label= GAP-3  
OTHER INFORMATION: /note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11420..11617

OTHER INFORMATION: /label= EXON-4  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11721..11722  
OTHER INFORMATION: /label= GAP-4  
OTHER INFORMATION: /note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 13354..13436  
OTHER INFORMATION: /label= EXON-5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 15044..15160  
OTHER INFORMATION: /label= EXON-6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 17245..17410  
OTHER INFORMATION: /label= EXON-7  
US-08-447-570-3  
Query Match 2.0%; Score 33.4; DB 1; Length 17410;  
Best Local Similarity 62.7%; Pred. No. 29;  
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 3 AATATAAACTCTCAGAAATAGAAATTAAGTATTCAGAGGCTAAATATGCAAAAACC 62  
DB 4401 AAAAAAAAAAAAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTAAAC 4342  
QY 63 GAAGAAACCTTTGGAAATTT 85  
DB 4341 AAGAAAGAAACAAAGAAATTT 4319  
RESULT 25  
US-08-449-700-3/C  
Sequence 3, Application US/08449700  
Patent No. 5863758  
GENERAL INFORMATION:  
APPLICANT: OPPERMAN, HERMANN  
APPLICANT: OZKAYNAK, ENGIN  
APPLICANT: KUBERASAMPATH, THANGAVEL  
APPLICANT: RUEGER, DAVID C.  
APPLICANT: PANG, ROY H. L.  
TITLE OF INVENTION: OSTEOGENIC DEVICES  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: 53 STATE STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: U.S.A.  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449, 700  
FILING DATE: 21-FEB-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 810, 560  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 827, 052  
FILING DATE: 28-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 660, 162  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 621,988  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,849  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 616,374  
FILING DATE: 21-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 600,024  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 599,543  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 579,865  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 569,920  
FILING DATE: 20-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 483,913  
FILING DATE: 22-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 422,613  
FILING DATE: 17-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 315,342  
FILING DATE: 23-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 232,630  
FILING DATE: 15-AUG-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 179,460  
FILING DATE: 08-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER, EDMUND R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: CRP-001CP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
TELEFAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17410 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3192..3730  
OTHER INFORMATION: //label= EXON-1  
OTHER INFORMATION: //note= "START CODON BEGINS AT POSITION 3313"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10413..10414  
OTHER INFORMATION: //label= GAP-1  
OTHER INFORMATION: //note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10696..10891  
OTHER INFORMATION: //label= EXON-2  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10960..10961  
OTHER INFORMATION: //label= GAP-2  
OTHER INFORMATION: //note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN

OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11059..11211  
OTHER INFORMATION: //label= EXON-3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11351..11352  
OTHER INFORMATION: //label= GAP-3  
OTHER INFORMATION: //note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11420..11617  
OTHER INFORMATION: //label= EXON-4  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11721..11722  
OTHER INFORMATION: //label= GAP-4  
OTHER INFORMATION: //note= "APPROXIMATELY \_\_\_\_\_ BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 13354..13436  
OTHER INFORMATION: //label= EXON-5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 15044..15160  
OTHER INFORMATION: //label= EXON-6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 17245..17410  
OTHER INFORMATION: //label= EXON-7  
US-08-449-700-3  
Query Match 2.0%; Score 33.4; DB 2; Length 17410;  
Best Local Similarity 62.7%; Pred. No. 29;  
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 3 AATAAAGAACTAGATACAAATTAAGTATTTTCAGAGGGTAATATATACAAAACC 62  
|| ||||| | ||| ||||| ||||| | || ||||| |||||  
Db 4401 AAAAAAAAAAAAAAAAAAGAAAGAAAGAAAAAAAAAGAAAAAAAAATAAC 4342  
OY 63 GAAGAAAGAACTTTTGGAAATT 85  
||||| | |||||  
Db 4341 AAAGAAAGAAAGAAAGAAATT 4319  
RESULT 26  
US-08-449-699A-3/C  
Sequence 3, Application US/08449699A  
Patent No. 5958441  
GENERAL INFORMATION:  
APPLICANT: OPPERMAN, HERMANN  
APPLICANT: OZKAYNAK, ENGIN  
APPLICANT: KUBERASAMPATH, THANGAVEL  
APPLICANT: RUEGER, DAVID C.  
APPLICANT: PANG, ROY H.L.  
TITLE OF INVENTION: ANTIBODIES TO OSTEOGENIC PROTEINS  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: 125 HIGH STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```

FEATURE:
NAME/KEY: exon
LOCATION: 15044..15160
OTHER INFORMATION: //label= EXON-6
FEATURE:
NAME/KEY: exon
LOCATION: 17245..17410
OTHER INFORMATION: //label= EXON-7
US-08-449-699A-3
Query Match      2.0%; Score 33.4; DB 2; Length 17410;
Best Local Similarity 62.7%; Pred. No. 29;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 3 AATAAAACCTAGTATGAGTAAGTATTTCAGAGGGTAAATATGACAACCAACC 62
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4401 AAAAAAAAAAAAAAGAACAAAAGAAAAGAAAAAAGAAAAAAGAAAAAATTAAC 4342
QY 63 GAAGAAAACCTTTGGAAATT 85
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4341 AAAGAAAGAAAAACAAGAAAT 4319

RESULT 27
US-08-486-343A-1/c
Sequence 1, Application US/08486343A
Patent No. 6071695
GENERAL INFORMATION:
APPLICANT: OKRAYNAK, ENGIN
APPLICANT: OPPERMAN, HERMANN
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING
TITLE OF INVENTION: MORPHOGENIC PROTEIN EXPRESSION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
ADDRESSEE: INC.
STREET: 45 SOUTH STREET
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 07148
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,343A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, Edmund R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-091CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)-248-7000
TELEFAX: (617)-248-7100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 17415 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: homo sapiens
US-08-486-343A-1
Query Match      2.0%; Score 33.4; DB 3; Length 17415;
Best Local Similarity 62.7%; Pred. No. 29;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 3 AATAAAACCTATGCAGTAAGTAATAAGTATTTCAGAGGGTAAATATGACAACCAACC 62

```





RESULT 31  
 US-08-327-494A-3  
 Sequence 3, Application US/08327494A  
 Patent No. 5527678  
 GENERAL INFORMATION:  
 APPLICANT: Blaser, Martin J.  
 APPLICANT: Tummuru, Murali K.R.  
 APPLICANT: Sharma, Smita A.  
 TITLE OF INVENTION: cagb and cagc Genes for H. pylori and  
 TITLE OF INVENTION: Related Methods and Compositions  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
 STREET: 127 Peachtree Street, Suite 1200  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 303-3  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25

```

1  RESULT 32
2  PCT-US95-13659-1
3  ; Sequence 1, Application PC/TUS9513659
4  ; GENERAL INFORMATION:
5  ; APPLICANT: Blaser, Martin J.
6  ; APPLICANT: Tumuturu, Murali K.R.
7  ; APPLICANT: Sharma, Smita A
8  ; TITLE OF INVENTION: cagb and cagc Genes for H. pylori and
9  ; TITLE OF INVENTION: Related Methods and Compositions
10 ; NUMBER OF SEQUENCES: 8
11 ; CORRESPONDENCE ADDRESS:
12 ; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
13 ; STREET: 127 Peachtree Street, Suite 1200
14 ; CITY: Atlanta
15 ; STATE: Georgia
16 ; COUNTRY: USA
17 ; ZIP: 303-3
18 ;
19 ; COMPUTER READABLE FORM:
20 ; MEDIUM TYPE: Floppy disk
21 ; COMPUTER: IBM PC compatible
22 ; OPERATING SYSTEM: PC-DOS/MS-DOS
23 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
24 ; CURRENT APPLICATION DATA:
25 ; APPLICATION NUMBER: PCT/US95/13659
26 ; FILING DATE:
27 ; CLASSIFICATION:
28 ; ATTORNEY/AGENT INFORMATION:
29 ; NAME: Spratt, Gwendolyn D.
30 ; REGISTRATION NUMBER: 36,016
31 ; REFERENCE/DOCKET NUMBER: 2200.029
32 ; TELECOMMUNICATION INFORMATION:
33 ; TELEPHONE: 404/688-0770
34 ; TELEFAX: 404/688-9880
35 ;

```







```
APPLICANT: Jiang, Yugu
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aljun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C16
CURRENT APPLICATION NUMBER: US/09/605,785
CURRENT FILING DATE: 2000-06-27
NUMBER OF SEQ ID NOS: 835
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 140
LENGTH: 200
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(200)
OTHER INFORMATION: n = A,T,C or G
US-09-605-785-140
```

```
Query Match          2.0%; Score 33; DB 4; Length 200;
Best Local Similarity 54.1%; Pred. No. 5.8;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
QY 150 TTCTTCTGATTAACATTAACTATACGGTGTACGGCATACAAAGACACTTATTGTG 209
      |||| ||| ||| |||| |||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13 TTCTGTTGTGTGATTTACTATAGGGGTTNGCTTNTCTAANATACTTTTCATTVA 72
QY 210 GGAGCTCCCTGGTCTGTGTCAGAGGCAATACCTTTCATCAAGTTTGGCTT 260
      || || || || || || || || || || || || || || || || || ||
DB 73 ACANCTTTTGTAAGTGTCAGGCTGCACCTTGGCTCATANAAATTATTGTTT 123
```

Search completed: January 21, 2003, 19:33:00  
Job time : 354 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 18:00:23 : Search time 1674 Seconds  
(without alignments)  
15837.522 Million cell updates/sec

Title: US-09-869-433-1  
Perfect score: 1637  
Sequence: 1 gaataaaactacacagaa.....gttgatccctactcgtcttt 1637

Scoring table: IDENTITY\_NUC  
Gapop 10.0', Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estemu:\*  
5: em\_estcov:\*  
6: em\_estipl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	332.2	20.3	1866	11	AY107360 Zea mays
2	226.4	13.8	966	10	BE574942
3	197	12.0	928	14	H55010 HH042a Sorgh
4	195.8	12.0	600	12	BG097617 EST462136
5	190.4	11.6	567	10	AW931058 EST356901
6	184.4	11.3	691	12	BG269363 LO-3415T3

7	181.4	11.1	558	14	B0790720	B0790720 E4555 Chi
8	178	10.9	645	10	AV926113	AV926113 AV926113
9	168.4	10.3	664	13	B1405734	B1405734 115G06 Ma
10	168.2	10.3	715	10	BE459811	BE459811 EST415103
11	165.6	10.1	568	12	BG463286	BG463286 EML_48_F1
12	165.2	10.1	602	14	B0754402	B0754402 EBCa01_SO
13	165.2	10.1	685	13	BM448586	BM448586 DSA025G04
14	163.8	10.0	563	14	BQ490006	BQ490006 26-E9427-
15	163.6	10.0	573	13	B1300874	B1300874 B1300874
16	163.2	10.0	527	9	AJ499428	AJ499428 AJ499428
17	163.2	10.0	529	13	B1271311	B1271311 NF051C03F
18	162.6	9.9	678	9	AJ432835	AJ432835 AJ432835
19	160.8	9.8	787	13	B1177811	B1177811 EST518756
20	158.4	9.7	518	10	AW738613	AW738613 EST340040
21	158	9.7	686	14	BQ255402	BQ255402 MTNAC62TK
22	157.8	9.6	637	13	BQ482169	BQ482169 BQ482169
23	157.4	9.6	624	12	BF460154	BF460154 066G08 Ma
24	155.8	9.5	567	14	BM954156	BM954156 sam70h06
25	155.6	9.5	609	13	BM322252	BM322252 PIC1_2_B0
26	154	9.4	532	14	BQ469843	BQ469843 HX01A14T
27	149.6	9.1	700	14	BQ147136	BQ147136 NF034F08F
28	148	9.0	499	12	BF459810	BF459810 065F05 Ma
29	148	9.0	567	14	BQ507193	BQ507193 EST614608
30	147.2	9.0	462	9	AU057209	AU057209 AU057209
31	146.8	9.0	529	13	BM357837	BM357837 12111-F7
32	146.2	8.9	469	14	C74932	C74932 C74932 Rice
33	146	8.9	581	10	AW924705	AW924705 WS1_71_D0
34	144.6	8.8	522	12	BG322472	BG322472 EML_28_D0
35	143.8	8.8	744	12	BF460035	BF460035 072A12 Ma
36	143.6	8.8	602	12	BF053092	BF053092 EST438322
37	143.6	8.8	667	10	AW509457	AW509457 s137d04.y
38	143	8.7	606	12	BF423978	BF423978 s502f04.y
39	142.8	8.7	663	12	BG456211	BG456211 NF075807F
40	141	8.6	575	12	BE726529	BE726529 894092F04
41	140.8	8.6	557	10	AW671962	AW671962 LG1_353_G
42	140.8	8.6	648	10	AW255797	AW255797 ML866 pep
43	138.8	8.5	601	10	AW255224	AW255224 ML219 pep
44	136.4	8.3	556	13	B1463735	B1463735 B1463735
45	136.2	8.3	477	9	AJ499380	AJ499380 AJ499380

## ALIGNMENTS

RESULT 1  
AY107360  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays.  
Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 1866)  
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.  
Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes  
Unpublished (2002)  
2 (bases 1 to 1866)  
Coe,E.C.  
Direct Submission  
Submitted (25-Apr-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA  
LOCATION/Qualifiers  
1.1866  
/organism="Zea mays"  
/db\_xref="MaizeDB:635637"  
/db\_xref="taxon:457"  
/clone="PC009845"

```
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
```

/note="this sequence is part of a project of ESR assemblies resulting from the application of public configs to seed Dupont colligis; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maizee Mapping Project"

BASE COUNT	437 a	446 c	467 g	516 t
ORIGIN				

Query Match	20.38;	Score 332.2;	DB 11;	Length 1866;
Best Local Similarity	55.0%;	Pred. No. 3.3e-67;		
Matches 725; Conservative	0;	Mismatches 578;	Indels 16;	Gaps 3;

QY	226	GTGGAGAGGCAATACCTTTCATCAAGTTTGGCTGTGGCCCTGTGCTATTAATCTTA	285
Db	18	GGCGCGAGATCATCCCTTCTCTCAAGACCTGGGTCAACCTGCCCATGGATCGGCGCTTCA	77
QY	286	TGCTTATTTATATGCAAAAGCTAAGTAATATTTTAAAGTAAAGAGGACCTTATTTATAGCAGTGG	345
Db	78	TGCTCTCTACACCAACAGCTGGCGGATGTCTCTCCAAAGGAGGGGCTCTTCAAGCGGCTCA	137
QY	346	GAAGCCCTTTTAAATTTCCTTTGGCCCTGTGTCCGCACTGAATTTATATTCGGCTAGCGCATG	405
Db	138	TCTTCCGTTTATTCGCCCTCTCTGGGGGCCCTTCCGCTACAGTGGCTCAACCCATGCGCGACG	197
QY	406	TTTTTATATCTACAGATTTTGTGACCGTTTACAGGCAATCTACCTACCTCCAGATTTGGTAG	465
Db	198	CCATTCACCCCACTGCGCTGGCGGATGCGCTCTCGGCTGTGCTCGGGGCCACACTCTCTCG	257
QY	466	GACTGTGTTCATCTTAAAGAACTGGACATTTTGCTCATTTTATGTACTTGTGCTGAACAT	525
Db	258	GACCCGTGGCATCTCTCGGCTGGAGGCTTCTGCTCTTACAGTCATGGCGAGCTCT	317
QY	526	GGGGAACCGTCATGCTATCTCTATATGTCTGGGGATTTTCTATGAAATTACAAAATCC	585
Db	318	GGGGCACTGTCGTCATATACGCTCTTTCTGGGGGTTTCCAATCAGATTACTACAGTTG	377
QY	586	ACGAAAGCAAGCGTTTCTACGCTCTTTGGGTAATGCGAGCTATATTTCTTACAGCTT	645
Db	378	AAGAGCGCAAGATCTACCCGCTGTGGGCTTGGGGCCATGTGGCCCTCATCTTCT	437
QY	646	CTGCTGTGCAATTTGTTGGGCTTCAAGTTGAGAGCTTCCGTTTCTGAAGGTAGATC	705
Db	438	CTGGGCGCAGGTGCAAAATTTTCTCAACATGAGAAATTTGGGTCCTGGAGTGATG	497
QY	706	CTTGGGGAATTC-TTTAGCTCTTTTGAATGCGTATGACATTTGTATCTGAGCTTGTCTT	764
Db	498	GTTGGCAATTTGTTGAAGGACATGATGAGATAGTGGTGTACTGGGCTGTGTCATC	557
QY	765	ATGCGCAGTACTAGCTGATCATTAAGAAAGTTTACGCGATCTCCGCTTATTAATCA	824
Db	558	ACCGGTATCTATTGGGAGTGAACAAGTTGTATTGATACT-----A-NC	605
QY	825	GAACAAATGCAAAAAGGGGAAAAAGGTGCTAAACCTTAAATGAATATGAAGATAGCTTC	884
Db	606	GTCGCCGTGTGATGGGGAAGAAAGAGCAACCAAGCTCAGCATGGGTGAGGATAG	665
QY	885	CTCATCTTATAGATCTCCTTATATTCTTTTATTAACCTCTTGGTATTGCTCATGTGT	944
Db	666	AAGGTTCTGTTGATCTCGTATGTGAGGGATCTTGCTACATTTGTTGTGCTTATGA	725
QY	945	ATTTGCATTAATTAATGCAAGGACTTTGGAAAAAGCACCTAAACCTGAATATCTTAT	1004
Db	726	ATAAGCTTAACTCTGTGCGAGGTGACATGGAATCAAAATTAAAGCACAGTTCCTCAAGC	785
QY	1005	ATGAATGACTATAGTAGTTCATGGGAACTTCTCTTCTGAGACTGGGCTAGATATCCGTA	1064
Db	786	CCGAATGAATATCTTCATTCATGGGCGATTTCTCAATCGCACCGCGATAG--CTACA	842
QY	1065	CTTATATCTATTTTGTGGTGAACGTCAATTCGTAATTTGGATGGTAACTGAGACC	1124

Db	843	TTTAAATATGATGTTGTTAGGAAGAGTATCCGACAAAGTTTGGGTGGGGAGTTGGACGT	902
Qy	1125	CTAGTCACCTCCCTGTCATGGTCTTCCCTAACAGTATCGTTTCTTGCGCTCTTGTATCTTT	1184
Db	903	ACGATACAGCCCTGGACAGTGTACTCTCCACAGAGATGGGTGGTCTTCTCACATGATTTTGGTTT	962
Qy	1185	AGAAACCAAGCTTTCGGGGCTGGCGGTATGTTGGTACAAATCCCTTCATGTCAGTCTGGT	1244
Db	963	GGTGAGCCATTTACTCTCTCTTTATGACCAAGTTTGGATATACACCTTTCCTTGGCGCAGTC	1022
Qy	1245	GTTTTCGAGCTATACAGATATCTCTTTTGGAAATCCACAAATATAGCCTCTCTTGACATCA	1304
Db	1023	TATGTTGGAGCAATGCAGAAACATTTTCACTAAGAGTGCAGAAATATACAGTCTGTATGATCTC	1082
Qy	1305	ACTAAAGAATGGCTATATCCCTCTTGACCAAGACCAAAAGTCAAAAGTAAAGGCTGCT	1364
Db	1083	TGCAAAGAAATGGCATACATTCCTTTGGATGAGGATATGAAGCTGAAGATGAAGCGGCT	1142
Qy	1365	ATTGATGTGATGGCCGCGCCCTTTCGGAAATATGAGAGCATTTATATCAACAAGGTTTG	1424
Db	1143	ATTGATGTTGTGTGCACACCCCTTTGGGGAATCTGGAAGTGTCTGATTCAGCAGATTCAAG	1202
Qy	1425	CTCGTTATCTGTGAAGATTTGGAGCTATGACCCCTTATCTTTCGAGTAGTCTTCTTTTC	1484
Db	1203	ATCCGTCAATTTGGTTCCTCTCGCAACTTCGACACCCTTACCTTGGGGGAATATCTCTGGTG	1262
Qy	1485	ATCATTTGCATTTTGGTGGTTCCTGCACATTAAGTTAAACAACATTTCTTACGCGACATC	1543
Db	1263	ATTGTCCTTGACAGGCTGGGTGCTGTGAAGAGTCCCTTCGACCTGCAGAGTCTCTTCCCTGGG	1321

RESULT 2					
LOCUS	BE574942				
DEFINITION	BE574942	966 bp	mRNA	linear	EST 07-DEC-2001
ACCESSION	D06	Triphysaria versicolor	root-tip, early DMB0-induced transcript		
VERSION	BE574942	cdna	library	Triphysaria versicolor	cdna, mRNA sequence.
KEYWORDS	BE574942.1	GI:12001272			
SOURCE	EST.				
ORGANISM	Triphysaria versicolor.				
REFERENCE	Triphysaria versicolor.				
AUTHORS	Eunaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;				
	Asceridae; euasterids I; lamiales; Orobanchaceae; Triphysaria.				
	1 (bases 1 to 966)				
JOURNAL	Matvienko, M., Torres, M.J. and Yoder, J.I.				
MEDLINE	Transcriptional responses to the hemiparasitic plant Triphysaria				
COMMENT	versicolor to host plant signals				
	Plant Physiol. 127 (1), 272-282 (2001)				
	21437952				
	Contact: Yoder, J.I.				
	John I. Yoder Research Lab, Dept. of Vegetable Crops				
	University of California at Davis				
	137 Assumundson Hall, One Shields Drive, Davis, CA 95616, USA				
	Tel: 530 752 1741				
	Fax: 530 752 9659				
	Email: jiyoder@davis.edu				
	length = 966 bp.				

FEATURES	SOURCE
Location/Qualifiers	1. .966
/organism="Triphysaria versicolor"	
/db_xref="taxon:64093"	
/clone_lib="Triphysaria versicolor root-tip, early DMSO-induced transcript cDNA library"	
/tissue_type="root-tips ~5mm length"	
/dev_stage="3-4 weeks growth"	
/lab_host="E. coli"	
/note="Vector: PCR2.1 TA Cloning System, Invitrogen, Carlsbad, CA; PCR-based suppression subtractive hybridization cDNA library"	
BASE COUNT	222 a 186 c 240 g 318 t
ORIGIN	



Query Match 13.8%; Score 226.4; DB 10; Length 966;  
 Best Local Similarity 53.7%; Pred. No. 1.8e-42;  
 Matches 532; Conservative 0; Mismatches 431; Indels 27; Gaps 2;

QY 298 CAAGCTAAGTAATTTTAAAGTAAAGAGCCCTTAATTTTATGCAAGTGGGAACCCCTTTT 357  
 DB 4 CTAAGTTGGCTAATGTTTGTGGAAGATGCACTTTCTACACTTTATGTTGGCATTTCA 63  
 OY 358 TAATTTTCCCTGCGCTCCGAGCTAATTTATCCGCTACGGCAGCTTATGATTCCTA 417  
 DB 64 TTGGCTTTTGGGAGCATTTGGGTTTGCTTATCTCTCAGTAATTTTTCACACCCCA 123  
 OY 418 CAGAATTTGCTGACCGTTTACAGCCATCTACCTCAAGATTTGCTAGACCTGTTGCCA 477  
 DB 124 CTGGCCTTGTGATTAAGCTTTTGAACGTGTGGGTCGAGGTTCTTGAGACACTTGCTA 183  
 OY 478 TCTTAAGAACTGACATTTGCTGATTTTATGTACTTCTGTAAGTATGGAAGAGCTCA 537  
 DB 184 TTTTGGAGATTTGGAGCTTTTGTGTTCTATGTAATGCTGACCTTTGGGCGACTGTCG 243  
 OY 538 TGCTATCTATGTTCTGAGGATTTGCTAATGAATTAACAATAATCCAGCAAGCAAGC 597  
 DB 244 TTGTTCTGTTTGTGTTGGGGTTTGGCAATCAGATACACAACTGACGACGACCAAGA 303  
 OY 598 GTTCTACGCTCTTTCCGTAATCGAGCTAATTTCTTACTAGCTTCTGTCGTGCA 657  
 DB 304 AATTCATACCTTTTGGGCTCGAGCAANTGTAGCACTTCTCTCGGCTGGGACTG 363  
 OY 658 TTTGTTGGGCTCAAGTTTGAGAGCTTCGTTCTGAAGTGTAGATCCTTGGGGAATTT 717  
 DB 364 TGAATATCTTTCCCAATGAGGCAAAATTTGGCCCGGCTGATGTTGGGCACT 423  
 OY 718 CTTTACGCTTTTGTGCTGCTATGATATGTAATGTAAGATAGCTTCTATCTTGTATA 777  
 DB 424 CGTTGAAGGGAATGATGATGATTTAGTGGGGAGTGGCTTATGATTTGTTCTATATT 483  
 OY 778 GGTGATCAATAGAACCTATTTGACCGATCTCGCTTATATATCCAGAAAGATGCAAA 837  
 DB 484 GGTGGTAAACAATTATGT-----TCCTCTCGAGCCGTA 519  
 OY 838 AGGGGAAAAAGTGTCTTAACCTTAATGAATGAAGATAGCTTCTATCTTGTATA 897  
 DB 520 GTATGAAAGAGAGAGAGAGCCGAAATGGAACGATGAGAGCTTAAAGTCTTGATAT 579  
 OY 898 GATGCTCTATATCTTTTAACTCTCTGTTATGCGATGATGCGATGATTTGCCATTAAT 957  
 DB 580 CAAAGAAATATATGAGATTTGGCCACTTGTGTGTCATATGCAATTAATGCAATTAAC 639  
 OY 958 TAATCGAAGTACTTGAAGATGCACTGAAGATGCAATATCTAATATGATGACTATA 1017  
 DB 640 TGTGTGAGGTTACATGAAATCAAAAGTCAAAAGTCACTTCCCAAGGCGCAAGGATAT 699  
 OY 1018 GTGAGTCTATGGGACCTTCTCTTCTGAGCTGCGGTGATATCGTACTTATCATGTAT 1077  
 DB 700 CTGCTTCATGAGTCACTTTTCACTACTACCGAATAGCAAACTTCAATATGATGCTTT 759  
 OY 1078 TTGTTGGTGAAGCTCAATGTAATTTGATGTTTACTGAGAGACCTAGTCACTCTG 1137  
 DB 760 TGAAGTCAATGATTTTAAAC---AAATAGCGTTGGGATGACACCAAGATTAAGCGCA 816  
 OY 1138 TCATGGTCTCTCTACAGATATCGTTTCTCTGCTCTGTTATCTTGAAGCAAGCTT 1197  
 DB 817 CCGTCTCTATACCTCTGGATTTGGCTTCTCTCTGTTGATTTCTTGCGTGAACCTTTTG 876  
 OY 1198 CTGGGCTGCTGATGTTGCGTACACCTCTCATCTAGCTGATGTTGTGCGAGCTA 1257  
 DB 877 CCCCCTGCTCGCGGCTTCGGAATGACTCCATATAGCAGCTGTATGTTGGGCGCA 936  
 OY 1258 TACAGAAATATCTTCGAAATCCCAAAAT 1287  
 DB 937 TGCAAAACATTTTATGTAAGAGCGCAAGT 966

RESULT 3  
 H55010  
 LOCUS  
 DEFINITION  
 H55010 Sorghum bicolor cv. TX430 Sorghum bicolor cDNA clone HH042  
 5' similar to adenine nucleotide translocator, chloroplast, mRNA  
 sequence.  
 accession  
 H55010  
 version  
 H55010.1  
 keywords  
 EST  
 source  
 sorghum.  
 organism  
 Sorghum bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoidae; Andropogoneae; Sorghum.  
 reference  
 1 (bases 1 to 928)  
 Wyrich, R., Dreesen, U., Brockmann, S., Streubel, M., Chang, C., Qiang  
 D., Paterson, A.H. and Westhoff, P.  
 The molecular basis of C4 photosynthesis in sorghum: isolation,  
 characterization and RFLP mapping of mesophyll- and  
 bundle-sheath-specific cDNAs obtained by differential screening  
 Plant Mol. Biol. 37 (2), 319-335 (1998)  
 journal  
 Plant Mol. Biol. 37 (2), 319-335 (1998)  
 comment  
 Contact: Westhoff P  
 Plant Molecular Biology  
 Heinrich-Heine-University  
 Universitaetsstrasse 1, D-40225 Dueseldorf, Germany  
 Tel: (49) 211-311-2338  
 Fax: (49) 211-311-4871  
 Email: west@uni-duessel.dorf.de  
 Seq primer: SK (17-mer).  
 features  
 source  
 1..928  
 /organism="Sorghum bicolor"  
 /strain="cv. TX430"  
 /db\_xref="taxon:4558"  
 /clone="HH042"  
 /clone\_lib="Sorghum bicolor cv. TX430"  
 /note="vector: Lambda ZAP II; leaf cDNA library,  
 unidirectionally cloned"  
 BASE COUNT 225 a 182 c 245 g 275 t 1 others  
 ORIGIN  
 Query Match 12.0%; Score 197; DB 14; Length 928;  
 Best Local Similarity 54.5%; Pred. No. 1.4e-35;  
 Matches 511; Conservative 0; Mismatches 405; Indels 21; Gaps 5;

QY 607 CTCTTTGCGATCGAGCTAATATTTCTTACTAGCTTCTGCTGTCGCAATGTTGGG 666  
 DB 1 CGCTGTTGGGCTTGCTCAATGAGCCCTACATCTCTGCGGCGACGGTGAATATT 60  
 OY 667 CTCAAGTTGAGAGCTTCGGTTCTGAAGGTGATGATCTTGGGGAATTTCTTACGTC 726  
 DB 61 TCTCAACATGAGGAAGATTTGGGCTCTCGAGTGAATGTTGGCAATTC----- 112  
 OY 727 TTTTATGCGATGACTAATTTTATCTGATCTGATGCTTATATGCGCGATTACGTCATCA 786  
 DB 113 -GTTAAAGCCATGAGCATAGTGTGTTGTTGCTTATATCATCACCGGTATCTATTGGG 171  
 OY 847 AAGTGTCTAATCAATTAATGAATGAAGATAGCTTCTCTATCTATGATGATCTCTT 906  
 DB 225 AGAAGACAAGCAAGCAAGCTGAGCATGAGGATGAAGATGAAGTCTGTTGATCTCCGT 284  
 OY 907 ATATCTTTTATTAATCTCTGTTGTTATGCTTATGCTATGATGATTAATCAAG 966  
 DB 285 ATGTAGGAGATCTTCCCATTTGGTCTTCTTATGATTAAGCATTAACCTTGTGGAGG 344  
 OY 967 TGACTTGGAAAAAGTCAAGTGAATGCAATATCTTAATGAATGATGATGATGATGATCA 1026  
 DB 345 TGACATGGAATCAAAATTTGAAGCAGAGT--TCCAAGCCCAAAATGAATATCTTATTTCA 403

OY	1027	TGGGGAACCTTCCTTCTGTGCAGGGGGTGATGTAACCGTACTATTTCATTTGTTGGTG	1086
OY	1087	GTAACGTCAATTCGTAATAATTTGGATGGATTACGTGAGACCCTTAGTCACTCCTGTCATGGTTC	1146
OY	404	TGGGCGATTTCCTCAACTGCCACCGCATAGCTAACCTTTACAATGATGTTGTTAAGGAGAG	463
Db	464	TAAATCTCA---GAAAGTTCCGGGTGGGAGTTGGACATCGANATCACACTCGAGTGTTC	520
OY	1147	TCCTAACAGGATTCGTTTTCTTCGCTCTTGTTATCTTTAGAACCACAGCTTGGGCGTG	1206
Db	521	TTCTACACAGAGATGGGTTCTTCCTACATGATTTTGTGGTGAGGCATTGACCTCCCTTA	580
OY	1207	TGCGATATCTTCGTTACAACTCCTTCATGCTGCTAGCTGTGGTTGCGAGTATACAGATA	1266
Db	581	TGACCAAGTTGGGATGACACCTTGTCTTGGGCGAGTCTATGTTGGAGCAATGCGAACA	640
OY	1267	TTCTTTGCAATTCACAAAATATCGCTCTCTTTGACTCAATGAAGAAATGGCCTATATCC	1326
Db	641	TTTTCATGATAGAGTGGATAAAGTACAGTCTGTGTTGATCTTGC AAAAGAGATGCATACATTC	700
OY	1327	CCTTCGACCAAGACAAAAAGTCAAAGGTAAGCGCTATGATGATGATGCGGCCCGCT	1386
Db	701	CTTTGGATGAGGATATGAAAGGTTAAAGGTAAGCAGCTATTGACCTGTGCTCAACCCCT	760
OY	1387	TGCGAAATTCAGAGAGAGAGCTTTTAATCCAAACAAGTTTGCTGCTTAATCTGTGAGAGATTATG	1446
Db	761	T-GGGAANTCTGAGAGTGGTTGATGTCAGACATTCATATCCTGTGCATTTGGATCTCTTG	819
OY	1447	GAGCTATGACCCCTTATCTTGCGATGATTTCTTTTCAATTCATTCGATTTGGTTGGTTT	1506
Db	820	CGAATTCAGACCCCCTACCTTGGTGGAAATACCTTCTGCTGATTTCTTGATGGCTGGGTG	879
OY	1507	CTGCAACTAAGTTTAACAACAATTCATTGAGCCAGTNC	1543
Db	880	CTGTAAAGTCCCTCGACATCACAGTTTTCACACCCTGGC	916
RESULT 4			
LOCUS	BG097617	600 bp	mRNA Linear EST_29-JAN-2001
DEFINITION	EST462136 potato leaves and petioles Solanum tuberosum cDNA clone		
ACCESSION	CGTB47J17 5' sequence, mRNA sequence.		
VERSION	BG097617		
KEYWORDS	BG097617.1 GI:12587652		
SOURCE	EST.		
ORGANISM	Solanum tuberosum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asterales; euasterids I; Solanales; Solanaceae; Solanum. 1 (bases 1 to 600)		
AUTHORS	van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J., Uutterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning ,C.M., Fry,W.E., Tanksley,S.D. and Baker,B. Generation of ESTs from potato leaves and petioles Unpublished (2000) Contact: Cathy Ronning The Institute for Genomic Research For clone info: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdnaresgen.com.		
JOURNAL COMMENT	Location/Qualifiers		
FEATURES	source		
	/organism="Solanum tuberosum"		
	/cultivar="Kennebec"		
	/db_xref="taxon:4113"		
	/clone="CGTB47J17"		
	/clone_lib="potato leaves and petioles"		
	/tissue_type="leaflets and petioles"		
	/dev_stage="8 weeks old plants"		
	/lab_host="SOLR"		
	/note=Vector: plusescript SK(-); Site_1: EcoRI; Site_2: XhoI; Tissue was supplied by Dr. Fry (Cornell University).		

[illegible]



Dd	418	TAAAGGGAAGACGCATTTGACGTGGTTCGCAACCOCCTTGGAANAATCCGAGAGCGCTCT	477
Oy	1409	AATCCAACAAGSTTTGCTCCTTAATCTGTGCGAAGTATTGGAGCTATGACCCCTTATCTTGC	1468
Dd	478	GATCCAGCAGTTCATGATCATCTTAACCTTTGGGACACTTGCCAAATTCAACCCCTTACCTTGG	537
Oy	1469	ACTGATTCCTCTTTTCATCATGTGATTTGGTTGGTTTCTGCAACGTAAGTTAAACAAT	1528
Dd	538	AGGAAATTCCTCTTGATTTGATTAAGTACGATCGATCGCTGGCAGCAGACGCTTTTGGACAATA	597
Oy	1529	ATTCTTAGCG	1538
Dd	598	GTTTACACAGG	607
RESULT 7	BQ790720	558 bp	mRNA linear EST 30-JUL-2002
LOCUS	BQ790720	E455 Chinese cabbage etiolated seedling library Brassica rapa	
DEFINITION	subsp. pekinensis cDNA clone E4555, mRNA sequence.		
ACCESSION	BQ790720		
VERSION	BQ790720.1	GI:22005682	
KEYWORDS	EST.		
SOURCE	Brassica rapa subsp. pekinensis.		
ORGANISM	Brassica rapa subsp. pekinensis. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica. Ryu,S.H., Yang,K.A., Lee,S.Y., Kim,H.-I., Cho,M.J. and Lim,C.O. Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA (2002)		
REFERENCE	Unpublished (2002)		
AUTHORS	Contact: Lim, C.O.		
TITLE	Plant Molecular Biology & Biotechnology Research Centre Gyeongsang National University #900 Gaze-dong, Jinju 660-701, Korea Tel.: 82 55 751 6255 Fax: 82 55 759 9363 Email: colim@nongae.gsnu.ac.kr Seq primer: T7.		
JOURNAL			
COMMENT			
FEATURES			
source	location/qualifiers		
	1..558		
	/organism="Brassica rapa subsp. pekinensis"		
	/cultivar="Jangwon"		
	/db_xref="taxon:51351"		
	/clone="B4555"		
	/clone_lib="Chinese cabbage etiolated seedling library"		
	/tissue_type="Etiolated seedling"		
	/lab_host="XL-1 Blue"		
	/note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"		
BASE COUNT	131 a 141 c 131 g 155 t		
ORIGIN			
Query Match	11.1%, Score 181.4; DB 14; Length 558;		
Best Local Similarity	59.7%; Pred. No. 6.2e-32;		
Matches 324; Conservative	0; Mismatches 216; Indels 3; Gaps 1;		
Oy	936	GCCATATGATTTGGCATTAACCTTAATCGAAGGACCTTGGAAGAAGCACCTGAAACTGAACTGCAA	995
Dd	13	GGTATGGAAATTTAGTATCAACCTTGTTAAGTCACATGGAATAATCAACCTTAACCTGAG	72
Oy	996	TATCCTATATGAATGATCATATAGTAGTTTCATGGGGAACCTTCTCTTGACACTGGCGTA	1055
Dd	73	TTCCCTAATGCCAAGAGTACTCAGCGCTTTATGGTGACTTCTCCACCTGCACGTGTATT	132
Oy	1056	GTATCCGACTATATCAGTCTATTTGGTTGGTGTAACGTCATTCGTAAATTTGGATGCTTA	1115
Dd	133	GCAAC--ATTCAACATGATGGTCTTAAGTCAATACGTATTCACACAGTATAGATGGGGA	189
Oy	1116	ACTGAGCCCTTAGTACGTCCTGTATGGTTCTTCCCTAACAGATATGTTTTCTTCGCTTT	1175
Dd	190	GTAGCTGGAGATGATCCCAACCTTCTGTATGACCGGAGTTGGCTTCTTCTCTG	249

QY	1176	GTATCTTTAGAAACCAACACTCTTGCGCGTGGCGCATATCTGGGTCAACTCCTCTCAG	1235
Db	250	ATATCTTTTGGTGGCCCGTTCGCACCACTGGTTGCCAACTTGGATGACACCCGCTCTC	309
QY	1236	CTAGCTGTGGTGTGTGAGACTATACAGAAATATCTTTTCGAAATTCACAAATATACGCTCTC	1295
Db	310	CGACGCTGTAGCTTGGTGGCTTTACAGAAATCTTTCAGCAAGAGCGCCAAAGTTACAGCTTGG	369
QY	1296	TTTGACTCACTAAAGAAATGGCCTATATCCCTCTTGGACCAAGACCAAAATGCAAAAGT	1355
Db	370	TTTCGATCTCTTGGCAAAAGAAATGGCTATATATCCATTGGATGAGAGACACCAAGGTTAAAGGC	429
QY	1356	AAGCGCTGATATGATAGTAGTATTCGCCGCCCTTCGAAATTCAGAGAGACCTTAATTCAA	1415
Db	430	AAAGCTGCTATTTGATGATGTTGTCTGCAACCACTTGGGAAATTCAGCGCGTCTCTGATCCAG	489
QY	1416	CAAGGTTTGCTGTTATCTGTGGAAATATTTGAGCTATGACCCCTTATCTTCAGTGAAT	1475
Db	490	CAGTTTATGATCTTCACGTTGGCTGCTACACTGGCAGCACTTACGTCCTTACCTTGGAGTCATTT	549
QY	1476	CTT 1478	
Db	550	CTT 552	
RESULT 8			
AV926113			
LOCUS	AV926113	645 bp	mRNA
DEFINITION	AV926113 K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA clone basd18c14 5', mRNA sequence.		
ACCESSION	AV926113		
VERSION	AV926113.1	GI:18221910	
KEYWORDS	EST.		
SOURCE	Hordeum vulgare subsp. vulgare.		
ORGANISM	Hordeum vulgare subsp. vulgare Euryotia, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poidea		
REFERENCE	1 (bases 1 to 645) Sato, K., Saitoh, D. and Takeda, K. Barley EST sequencing project in NIG and Okayama Univ Unpublished (2002)		
AUTHORS	Contact: Tadasu Shino-i		
TITLE	National Institute of Genetics		
JOURNAL	1111 Yata, Mishima, Shizuoka 411-8540, Japan		
COMMENT	Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshino@genes.nig.ac.jp.		
FEATURES	Location/Qualifiers		
Source	1..645		
	/organism="Hordeum vulgare subsp. vulgare"		
	/cultivar="Haruna Nijo"		
	/db_xref="taxon:112509"		
	/clone="basd18c14"		
	/clone_lib="K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves"		
	/tissue_type="seedling leaves"		
	/dev_stage="second leaf stage"		
BASE COUNT	108 a	202 c	168 g 167 t
ORIGIN			
Query Match	10.9%	Score 178;	DB 10; Length 645;
Best Local Similarity	54.8%;	Pred. No. 3.9e-31;	
Matches 352;	Conservative 0;	Mismatches 290;	Indels 0; Gaps 0;
QY	123	AAGAAAGTTGGCCAAATTTCTCCTATATGTTCTTCTTCCTTATACATTTAACTATACGGTGA	182
Db	2	AAGAAGATCTGAGCCCTTGCTGCTCAATGTTCTTCTTCATCTCTTCACTACACCACTCCG	61
QY	183	CGCGATACAAAGACACTCTATTTTGGGAGCTCGCTGTTCTGTTGTCAGAGGCAATTACT	242

```

Db      62  AGGACACCAAGAGCTGCTGCTGACCGCAAGGAGCAGCGCCAGATCATCCCT 121
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      243  TTGATCAAGTTTGGCTTTGCTCCCTGCTATTTATCTTTATGTAAGAAG 302
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      122  TTGCTCAAGACGTGGGTCAACCTCCCATGAGGCTGCTGCTTCTTACTTCAAG 181
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      303  CTAAGTAATATTTTAAAGTAAGCAGGCTTATTTATGACGGGAACGCTTTTAAAT 362
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      182  CTGCGGAGCTGCTCCAGAGAGGCGCTTCTGACCGCATCTTCCCTTCATCGCC 241
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      363  TTTTGGCCCTGCTCCGACTGTAATTTATCCGCTACGCGATGTTTTACATCTACAGAA 422
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      242  TTTCTGGGCTCTCGGATAGCTGCTACCTATGCGGAGCAGCATCCAGCCAGCGCG 301
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      423  TTTGCTGACCGTTTACAGGCCATCTACCTCCAGAGATTGCTAGACTGTTCCATCTTA 482
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      302  CTGCGGAGCGGCTCTCCGCTGCGCGCCAGCTTCTGGGCGCCGCTCCATCCTTA 361
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      483  AGAAGCTGACATTTGCTGCTCATTTATGTAAGTATGGAAGTATGGAAGCGTCAATGCTA 542
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      362  CCGCTGAGAGTTTCTGCTCTTTTACGTTATGCGGAGCTCTGGGCGAGCGTGTGCTATC 421
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      543  TCTTAATGTTCTGGGATTTGCTAATGAATTAACAAATCCAGCAAGCAAGCGTTTC 602
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      422  TCCGTCGCTGTTGGGATTTGCCAATGATTAACAGCGTTGAGAGGCTTAAGAGTTT 481
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      603  TAGCGCTCTTTGGTATCGGAGCTAATTTCTTACTAGCTTCTGCTGCTCAATTTGTT 662
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      482  TACCACATCTTGGGCTTGGGCGCAATGCGCTCATTTCTGCTGCGCAGCGGAGAA 541
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      663  TGGGCTTCAAGTTGAGACTTCCGTTTCTGAAGGTGTGATCCCTGGGGAATTTCTTA 722
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      542  TACTTCTCAAAATGAGGAGAAATTTGGGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 601
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      723  GCTCTTGTGATGCTATGACTATTTATCTGAGACTTGTCTT 764
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      602  AAGGAGATGATGAGCATAGTGTGCTGGGTTGCTGCTATT 643
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
BI405734
LOCUS      115606 Mature tuber lambda ZAP Solanum tuberosum cDNA, mRNA
DEFINITION
Sequence.
ACCESSION BI405734
VERSION   BI405734.1 GI:15185148
KEYWORDS EST.
SOURCE    potato.
ORGANISM Solanum tuberosum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE EST415103 tomato developing/immature green fruit Lycopersicon
AUTHORS   Nielsen,K.L., Crookshanks,M., Emmersen,J. and Welinder,K.G.
TITLE     EST-sequencing of mature potato tuber (var. Kuras)
JOURNAL   Unpublished (2000)
COMMENT   Unpublished (2000)
           Contact: Karen G. Welinder
           Institut for bioteknologi
           Aalborg Universitet
           Sohngaardsholmsvej 49, 9000 Aalborg, Denmark
           Tel: +45 96358467
           Fax: +45 98141808
           Email: kwebio.auc.dk
           High quality sequence stop: 664
           POLYA-No.

```

```

FEATURES
SOURCE      1..664
             location/Qualifiers
             /organism="Solanum tuberosum"
             /cultivar="Field grown Kuras"
             /db_xref="taxon:4113"
             /clone_lib="Mature tuber lambda ZAP"

```

```

/tissue_type="tuber"
/note="Vector: lambda ZAP"
BASE COUNT      171 a      143 c      157 g      193 t
ORIGIN

```

```

Query Match      10.3%; Score 168.4; DB 13; Length 664;
Best Local Similarity 56.4%; Pred. No. 6.8e-29;
Matches 335; Conservative 0; Mismatches 256; Indels 3; Gaps 1;

```

```

Qy      939  TATGTAATTTGATTAATTAATCAAGTACTGTAAGAAAGTCAAGTCAATAT 998
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1  TATGATTAATTAATTAATCAAGTACTGTAAGAAAGTCAAGTCAATAT 60
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      999  CCTAATATCAATGACTATAGTATGATTCATGGGAACTTCTCTTGAGCTGGCGTAGTA 1058
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      61  CCAAGCCCAAGTAATATCTCATATTCATGAGGATGAGTCAAGTCAATGTAATGTA 120
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1059  TCCGATTAATATATTCATATTTGTTGTTGTAAGCATTCGTAATTTGGATGTTAACT 1118
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      121  ACTTTCATATATGATGTTGTTAGTCAATGCA---TTTTCAGACAGATAGGTTGGGAGCA 177
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1119  GGAGCCCAAGTACATCTGCTGATGTTCTCTAAGAGATATGTTTCTGCTCTTGT 1178
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      178  GCAGCCAAATTAACACCTACAGCTTCTGCTTAACCGGAGTTGTTTCTCCCTGCTT 237
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1179  ATCTTTAGAACCAAGCACTTGGGCTGCTGCTATGTTGGGTACAACTCTCTCATGCTA 1238
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      238  TTGTTTGTGCCCCCTCTACACCTACTCTTGGAGTTGGAAATGATGCTCTCTTACGA 297
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1239  GCTGTTGTTGTTGGAGCTTACAGATATTTCTTGAATTCACAAATATACGCTCTCTT 1298
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      298  GCTGTTATATGTTGGTATGATGCAATGCAATTTTCAGTAAGAGCAAGTATGTTT 357
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1299  GACTCAACTAAGAAATGCGCTATATCTCTTGAACCAAGACCAAAAGTCAAGTAG 1358
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      358  GACCCCTGCAAGAAATGCGCTATATCTTGGATGAGACCAACAGGTTAAAGGAAG 417
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1359  GGTGATATGATGTTGTTCCCGCCGCTCGGAAAAATCGAGAGAGGAGGAGGAGGAGGAGGAGG 1418
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      418  GCAGCAATGATGTTGTTGCAATCCATCGGAAGTGTGGAGAGCTTGTATACAACTA 477
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1419  GCTTGTCTGTTATCTGTGAAGTATTTGAGGAGTATGAGCCCTTATCTTGCAGTATCTT 1478
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      478  TTGATGATTTTGTACTTTTGTTCACCTTGGCAGCTGACACCTACTCGGCGGTGCTC 537
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1479  CTTTTCATCATGCTATTTGTTGTTGTTCTGCAACTAAGTAAACAACATATTC 1532
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      538  TTAGTATATGTTCTTCCATGCTTGGGAGCAGCCAAAGTCTTGGATGACAGATTTC 591
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
BE459811
LOCUS      715 bp      mRNA      linear      EST 18-MAY-2001
DEFINITION EST415103 tomato developing/immature green fruit Lycopersicon
AUTHORS   Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Romling,C.M.,
           Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
           S.D.
TITLE     Generation of ESTs from tomato fruit tissue, immature green
JOURNAL   Unpublished (2000)
COMMENT   Contact: CUGI
           Clemson University Genomics Institute

```



Dd	178	TTGGGACTGGCGTGGTTCTTCCTCCTGCATTTTGTTGGGCAACCAFTTACTCOCCTG	237
Oy	1206	GTCGCTATGTTCGGTAGACAACCTCCCTCATGCTAGCTGTGGTTGTGCGAGCTATACAGATT	1265
Dd	238	CTTGCGACGATGGGATGATACCCTCTCTCTTGCGGGCTGTTTATGTGGGGTGCCTTCAGAAC	297
Oy	1266	ATTCTTTTGGMAATCCACAATAATAGCCTCTCTTTTGACTCTAACTAAGAATAATGGCTATATC	1325
Dd	298	ATATTTACCAAGAGTGAAGATACGTTTATTCGCCCGTCCAAGAAATATGGCTATATT	357
Oy	1326	CCTCTTGACCAAGACCAAAAAGTCAAAGGTAGCGCTGCTATGTAGTGTATGGCCGCCGC	1385
Dd	358	CCTTTGATGTGAGATATATAGATTGAAGGAAAAGCAGCCATTGATGTGTCTGCAACCCA	417
Oy	1386	TTCGGAATATCAGAGAGAGCTTTTATCCACAAGTTTGCTGCTTATCTGTGAAGATT	1445
Dd	418	TTGGGCAAGTCCGGTGGTGGCGGTATCCACAGTTTCATGATCCGATATTTGGGCTTTTG	477
Oy	1446	GCAGCTATGACCCCTTATCTTCGACAGTATCTCTTTTCATCATATTCATTMTGGTGGTT	1505
Dd	478	GCAACCTGACGCCCCCTACCTTGCGGTACTGTGTGTCATCGTCGTCATAGCTAGCT	537
Oy	1506	TCTGCAACTAGTATAACAACATATCTTA	1535
Dd	538	GCGGCAAGATCATTTGGACAAGCATTCYCA	567
RESULT_13			
BM448586			
LOCUS	BM448586	685 bp	mRNA linear EST 01-Apr-2002
DEFINITION	DS4025G04.59793 An expressed sequence tag database for the halotolerant green alga, Dunalieella salina Dunalieella salina cDNA clone DS4025G04 5, mRNA sequence.		
ACCESSION	BM448586		
VERSION	BM448586.1 GI:19854158		
KEYWORDS	EST.		
SOURCE	Dunalieella salina.		
ORGANISM	Dunalieella salina Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Dunalieellaceae; Dunalieella. 1 (bases 1 to 685)		
REFERENCE	Cushman,J.C.		
AUTHORS	An expressed sequence tag database for the halotolerant green alga, Dunalieella salina		
TITLE	Unpublished (2002) Contact: Cushman JC Department of Biochemistry University of Nevada MS200, Reno, NV 89557-0014, USA Tel: 775-784-1918 Fax: 775-784-1650 Email: jcushman@unr.edu		
JOURNAL	PCR Primers		
COMMENT	FORWARD: T3 20mer BACKWARD: T7 21mer Plate: 025 row: G column: 04 Seq primer: T3 20mer High quality sequence stop: 685.		
FEATURES	Location/Qualifiers		
SOURCE	1..685		
	/organism="Dunalieella salina"		
	/db_xref="taxon:3046"		
	/clone="DS4025G04"		
	/clone_lib="An expressed sequence tag database for the halotolerant green alga, Dunalieella salina"		
	/tissue_type="Cells, which was adapted in 2.5M NaCl via a incremental series from 1.7 to 2.0 to 2.25 to 2.5 M NaCl, were exposed to 3.4 M NaCl for 5 hours"		
	/cell_type="Green"		
	/note="Vector: lambda Uni-zap XR. Bluescript SK-; Site_1: EcoRI; Site_2: XhoI. Library construction was performed according to Stratagene's recommended protocol for the lambda Uni-zapXR vector and cDNA synthesis kit."		













QY	864	ATGATATGAAAGATAGCTTCCTCATCTGTATGATGATCCCTTATATCTTTATTA	CT	923
Db	349	ATGACCACAATGAGAGCTTGAAGTCTTGCTCTTCAAAATATATCAGGATCTTGCC		408
QY	924	CTCTTGTTATTCCTTCATATGATATTTGGCATTTAACCTTAATCGAAGTAC--TTG	GAAGTCA	982
Db	409	ACATATGGTTTGAAGCATATATGGCATTTAGTATCAACCTTGTGAAGTATACAT	GGGAAGTCAAA	468
QY	983	GCTGAACCTGCAATTTCTTAATATGAATGACATATAGTATGATCATGGGAACTTCTCTT		1042
Db	469	GCTCAAAAGCTCAGTTCCTCCAAAGCCCAATGAATATCTCATTTCAATGAGGAGCTTCTCAAC		528
QY	1043	CTGAGCTGGCGTACATTCCTGACTATATATGCTATTTTGTGGTGAACGTCATCTGTAA		1102
Db	529	TGCTACTGTGAATAGCAACTTTTCACATATGATGTTGTTAAGTCAATGGA---TTT	TTTGACAA	585
QY	1103	ATTGGATGCTTTAACTGAGCCCTAGTCACTCTGTCAATGTTCTCTTAACAGATTCGT		1162
Db	586	GTAATGGGAGGAGACGACGACCAAGATTAACATCTTCAAGTCTTGCTTAACCGGAGTTGG		645
QY	1163	TTTTCTTCCTCTTGTATTCCTTTAGAAACCAAGCTTCTGGGCTGCGCGGTATGTGGTAC		1222
Db	646	TTTTCTTTCCTGCTTTTGTGTTGGTGCCCTTAGCACCCTACTCTTGGGAAGTTTGGAA	T	705
QY	1223	AACTCTTCATGCTAGCTGTGGTGTGGAGAGCTTATACAGATATTTCTTGGAAATCCAC		1282
Db	706	GACCTCTCTTCAGACGCTGTCTATGTGGGAGCAATGACAAATTTTCAGTAAGAGTGC		765
QY	1283	AAAAATA 1288		
Db	766	AAAGTA 771		

RESULT	20
LOCUS	AM738613
DEFINITION	AM738613 518 bp mRNA linear EST 18-MAY-2001
ACCESSION	Lycopersicon esculentum flower buds, anthesis, Cornell University
VERSION	Lycopersicon esculentum cDNA clone cTODN14 5', mRNA sequence.
KEYWORDS	AM738613
SOURCE	AM738613.1 GI:7647558
ORGANISM	EST. tomato. Lycopersicon esculentum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 518) van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang, 'F.', Hansen,T.S., Craven,M.B., Bowman,C.L., Rønning,C.M., Nierman ,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D. Generation of ESTs from tomato flower tissue, anthesis Unpublished (1999) Contact: CUGI Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA Email: <a href="http://www.genome.clemson.edu/orders/index.html">http://www.genome.clemson.edu/orders/index.html</a> 5 prime sequence.
TITLE	JOURNAL
COMMENT	

FEATURES	Location/Qualifiers
source	1. .518

```

/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CTOD7N14"
/clone_lib="tomato flower buds, anthesis, Cornell
university"
/lssue_type="flower"
/dev_stage="anthesis"
/ncbi_vector="pBluescript SK(-); Site.1: EcoRI; Site.2:
XhoI; supplier: Tanksley; flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA496). They

```

we were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

Query Match	9.7%	Score 158.4	DB 10	Length 518
Best Local Similarity	58.2%	Pred. No. 1.5e-26		
Matches 298	Conservative 0	Mismatches 211	Indels 3	Gaps 1

QY	927	TTGGTTATTGGCTATGGTATTTGGATTAACCTTATGCAACTGGATCTGGAATACTGACGTG	986
Db	4	TTGGTTGTAGCATATGGCATTAGATATAACCTTGTGAAGTTACATGGAAGTCAAAAGCTC	63
QY	987	AAACTCAATATCTTAATATGAATAGACTATAGTAGTTCATGGGGAACCTTCCTCTCTGG	104
Db	64	AAACCTCAGTTCCCAACCCCCAAGAAATCTGCTCATTTCAATGGGTGACTTCTCAACTGCT	123
QY	1047	ACTGGCGAGATCCGACTTATTCATCATCTATTGTGGTGCTGTACGCAATCTGTAATTT	110
Db	124	ACTGGATAGCAACTTTCACAAATGATGTTTGTAAATGATGA---TTTTCGACAAAGTAC	180
QY	1107	GGATGGTTAACTGGAGCCCTAGTACTCTGTGCATGGTTCTCCCTAACAGATCGTTTTC	116
Db	181	GGATGGGAGACAGACGCCAAGATAACACTACAGTCCCTGCTCTACCGGAGTGGTTTC	240
QY	1167	TTGCGCTTGTTATCTTTAAAAACCAAGCTTTCGGGCTGGTGGCTATGTTGGGTACAACT	122
Db	241	TTTCCTCTGTTTGTGTTGGTGCCCTCTTAGCACACTACTCTTCGAAAGTTTGGAAATGACT	300
QY	1227	CCTCTCAAGCTAGCTGTGTTGTTCGAGACTATACAGAAATATCTTTTGGAAATCCACAAA	128
Db	301	CCCTTTCTTACAGAGCTGCTATGTGGGTGCAATGCAGAACATTTTCAGTAAAGATGCAGA	360
QY	1287	TAGCGCTCTTGTACTCACTAAAGAAATGGCCCTATATCCCTTTACCAAGAGCAAAA	134
Db	361	TATAGTTTGTTCACCCCTCGCAAGAAATGGCTACTATTCCTTTGGATGAAGACACCAAG	420
QY	1347	GTCAAAGCTAAGCTGCTATTGTATGTAGTTGACCGCCGCTTTCGAAATACAGGAGAGCT	140
Db	421	GTTTAAAGGAAGGACGCAATCGATGTGTGTGCAAATCCACTGGGAAAGCTGTGAGAGAGCT	480
QY	1407	TTATATCAACAAGTTTGCATCTTATCTGAG	1438
Db	481	TTGATACAGCAATTCATGATTTTGTGACTTTTGG	512

RESULT 21	
BQ255402	
LOCUS	BQ255402
DEFINITION	BQ255402 686 bp mRNA linear EST 06-MAY-2002
ACCESSION	MFMNACC27TM KVKC Medicago truncatula cDNA clone pKVKC-3FZ, mRNA sequence.
VERSION	BQ255402
KEYWORDS	BQ255402.1 GI:20456155
SOURCE	EST.
ORGANISM	bareil medic. Medicago truncatula Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE	1 (bases 1 to 686)
AUTHORS	Vandenbosch,K., Endre,G., Silverstein,K., Town,C.D., Van Aken,S., Uttekatch,T., Cheung,F., Frazer,C.M., Lambin,A.-F. and Retzel,E.F.
TITLE	The Medicago truncatula 'kliloclone' set: A selection of ESTs from various libraries rearrayed for gene expression studies
JOURNAL	unpublished (2002)
COMMENT	Contact: Vandenbosch K

Department of Plant Biology  
University of Minnesota  
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA  
Tel: 612 624 2755  
Fax: 612 625 1738  
Email: kvandenb@cbs.umn.edu

Minnesota sequence name: MTN62TRM. Alias clone name: KV2-1D1.  
This sequence represents the 5' end of the cDNA insert. More  
information is available at: www.medicago.org  
Seq primer: 5'ggc tct aga act agt gga tc 3'.

## FEATURES

source

1. 686

/organism="Medicago truncatula"  
/cultivar="A17"  
/db\_xref="taxon:3880"  
/clone="pkvkc-3f2"  
/clone\_lib="KVKC"  
/tissue\_type="mixed tissues"  
/dev\_stage="various stages"  
/lab\_host="XLOLR"  
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:  
XhoI. cDNA was prepared from polyA+ enriched RNA. The cDNA  
was directionally ligated into the UniZap XR vector from  
Stratagene and packaged using Gigapack III Gold packaging  
extracts. Plasmids containing cDNA inserts were excised  
from the recombinant lambda-Zap phage using Ex-assist  
helper phage and propagated in XLOLR cells."

BASE COUNT 171 a 148 c 155 g 188 t 24 others  
ORIGIN

Query Match 9.7%; Score 158; DB 14; Length 686;  
Best Local Similarity 54.9%; Pred. No. 1.8e-26;  
Matches 333; Conservative 0; Mismatches 271; Indels 3; Gaps 1;

OY 981 CAGCTGAACGTCATATCTATATGATGACTATAGTTCATGGGAACTTCTCC 1040  
DB 10 CAGGTCAAGCTCAGTTCCTAGCCCAAGAGTACTCTTCTTATGGGAGACTCTCG 69  
OY 1041 TTCTGAGCTGGGAGTATCCGCTACTATATCTTGTGGTGGTACGTCATTCGT 1100  
DB 70 ACCGCAACTGG--AAATGCTACATTCATATGATGCTTCTAAGTCAATATATATTGAC 126  
OY 1101 AATTTGATGCTTACTAGGAGCCCTAGTCTCTGTCATGGTCTCTTACAGGATTC 1160  
DB 127 AATATGATGGGAGGAGTCTGCAAGATCAGACCTACAGTCTCTTTCGACGGAGTT 186  
OY 1161 GTTTCTTCTGCTGTATATCTTTAGAAACCAAGCTTGGGCTGCTATGTTGGT 1220  
DB 187 GGTTCCTCTCTCATATGTTGGAGGCTCCAAATGACCGGCTCTGATCAATTTGA 246  
OY 1221 ACAATCTCTGATGCTACTGCTGCTGTCGAGCATATCAAAATTTCTTCGAAATCC 1280  
DB 247 ATGACTCCACTGACAGCTGTATATGTTGGCCATGCAACATTTTTCAGCAAGAT 306  
OY 1281 ACAAAATAGCTCTCTTGTACTCACTAAGAAATGCTATATCCCTTTCAGCAAG 1340  
DB 307 GGTAGTACAGTTATGACCCCTGCAAGAAATGCTTACATTTCCCTTGTAGTAAG 366  
OY 1341 CAAAAATCAAGTAAAGGCTGCTATGATGATGTCGCCCGCTTGGGAAATCAGA 1400  
DB 367 ACTAAGTCAAGGGGAAACAGCCATTTGATGTTGGACCCATTTGGGAAATCTGA 426  
OY 1401 GGAGCTTATGCAACAAGGTTGCTGCTATCTGTCGAGATTTGGAGCTATGACCCCT 1460  
DB 427 GGAGCCCTTATCAAGTCAATGATCTTATCTTGGTCACTAGCAATCAATCA 486  
OY 1461 TATCTTGAAGTATCTTCTTTCATCTGCTATTTGGTGGTTCCTCACTAAGTGA 1520  
DB 487 TACCTTGGAGGCTGCTGTTGATGATTTGCTTGGCTGGTGGAGCAGCAAGTCTTGG 546  
OY 1521 AACAAATATCTTACGAGTCTGCTTTAAGAAACAAGAGTGGCTCAAGAAGATTGA 1580  
DB 547 GATACACAGTTATGCTATTTGGCTCAGAGAAAGACTGAGNAACAGATGGAAGAGCA 606  
OY 1581 GCTCTG 1587  
DB 607 GCAGCTG 613

RESULT 22  
BU482169 637 bp mRNA linear EST 23-MAY-2002  
LOCUS BU482169 K Sato unpublished cDNA library, strain H602 adult.  
DEFINITION heading stage top three leaves Hordeum vulgare subsp. spontaneum  
cDNA clone bah63ml9 5', mRNA sequence.

ACCESSION BU482169  
VERSION BU482169  
KEYWORDS GI:21160630  
SOURCE EST  
ORGANISM Hordeum vulgare subsp. spontaneum.  
Hordeum vulgare subsp. spontaneum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
; Triliceae; Hordeum.  
1 (bases 1 to 637)

REFERENCE Sato, K., Saitoh, D. and Takeda, K.  
AUTHORS Barley EST sequencing project in NIG and Okayama Univ  
TITLE Unpublished (2002)  
JOURNAL Contact: Tadasi Shin-I  
COMMENT National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES 1. 637  
source location/Qualifiers

/organism="Hordeum vulgare subsp. spontaneum"  
/strain="H602"  
/db\_xref="taxon:77009"  
/clone="bah63ml9"  
/clone\_lib="K. Sato unpublished cDNA library, strain H602  
adult, heading stage top three leaves"  
/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage" 1 others

BASE COUNT 90 a 238 c 164 g 144 t 1 others  
ORIGIN

Query Match 9.6%; Score 157.8; DB 13; Length 637;  
Best Local Similarity 56.6%; Pred. No. 2e-26;  
Matches 291; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

OY 112 CTCACGAGTAAAGATTCGCAATGTTCTTATGTTCTCTGATTAATTAATTA 171  
DB 123 CGCTACGCTCAAGAAATGCTGTCCTGCTGCTATGTTCTTCTATCCTTTCAACT 182  
OY 172 ATACGGTGTATACGCGATACAAAGACATCTTATTTGTGGAGCTCCTGTTCTGTCAG 231  
DB 183 ACACATCTGTAGAGACACCAAGAGAGCTGCTGCTGTCACCGCCAGAGGAGCAGCGCG 242  
OY 232 AGGCATTAATTTATCAAGTTTGGCTGTTGCTCCCTGTGCTATATCTTATGCTTA 291  
DB 243 AGATCATCCCTTCCCAAGACGTGGGTCAACCTCCCATGCGCATTCGCTATCTTTC 302  
OY 292 TTATGCAAGAGTAAATATTTTAAAGTAAAGAGGCTATTTTATGAGTGGGAACG 351  
DB 303 TCTACTCCAGCTGCGGAGCTCTCTCAAGAGGCGCTTCTTCAACCGTATCTTCC 362  
OY 352 CTTTTTAATTTCTTGGCCCTTCCGACGCTGTAATTTATCCGATACGGAGATTTTAC 411  
DB 363 CTTTATCCGCTTCTGCGGCTTCTGAGTACTGCTATGCTTATGACCGGAGCATCTCC 422  
OY 412 ATCTTACAGATTTGCTGACCGCTTACAGGCACTCTTACCTCAGAGATTTCTAGACTG 471  
DB 423 ACCCCAGGCGCTGCGGAGCGGCTCTGCGCTGCGGCTGCGGCGCCACCTTCTGGGCGCG 482  
OY 472 TTGCAATTAAGAACTTGACATTTTATGATTTTATGATTTTATGATTTTATGATTTT 531  
DB 483 TCGCATCTCAGCGCTGAGATTTCTGCTCTTCTTACGTTATGCGGAGAGCTTGGGGCA 542  
OY 532 GCGTCATGCTATCTGATGTTCTGAGGATTTGCTATGATAAATTTACAAAATTCACGAAG 591  
DB 543 GCGTGTCAATCTGCTGCTGTTTGGGAGATTTCCCAATCAGATTTACACAGGTTGAAGAG 602







RESULT	26
LOCUS	BQ469843
DEFINITION	HX01A14T HX Hordeum vulgare cDNA clone HX01A14 5-PRIME, mRNA sequence.
ACCESSION	BQ469843
VERSION	BQ469843.1
KEYWORDS	GI:21277625
SOURCE	EST.
ORGANISM	Hordeum vulgare. Hordeum vulgare Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae   Triliceae; Hordeum. 1 (bases 1 to 532)
REFERENCE	Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A. EST sequencing and analysis in barley (2002) Unpublished (2002) Contact: Steen Nils Molecular Markers Group, Department Genbank Institute of Plant Genetics and Crop Plant Research (IPK) Corrensstr. 3, 06466, Gatersleben, Germany Tel.: 039482-5522 Fax: 039482-5595 Email: stein@ipk-gatersleben.de Insert length: 532 Std Error: 0.00 Plate: 1 row: A column: 14 Seq primer: T3.
FEATURES	Location/Qualifiers 1..532 /organism="Hordeum vulgare" /cultivar="Barke" /db_xref="taxon:4513" /clone="HX01A14" /clone_id="HX" /tissue_type="apex" /dev_stage="apex (3-5 mm in size)" /lab_host="XL10-Gold"
BASE COUNT	95 a 155 c 136 g 146 t
ORIGIN	
Query Match	9.4% Score 154; DB 14; Length 532;
Best Local Similarity	56.2%; Pred. NO. 1.6e-25;
Matches 289; Conservative	0; Mismatches 225; Indels 0; Gaps 0;
OY	226 GTGAGAGCAATACCTTCATCAGATTGTGGCTTGTCGCCCTGACTATTATCTTTA 285
Dd	13 GCGGGGAATATCATCCCTTCCTCAAGACTGGGTCAACTGGCCATGGGCATCGGCTTA 72
OY	286 TGCTTATTATGCAAGCTAAAGTAATATTATAAGTAACGAGCCTTATTATTATGACGTG 345
Dd	73 TGCTGGCATACTCCAAGCTCTCCGACGTCCTCCGCGGAGGCGCTCTTCTACACCGTCA 132
OY	346 GAAGCCCTTTTAAATTTCTTTGGCCCTGTCGCCACGTATTAATATCCGCTAGCGCATG 405
Dd	133 TATTGCCCTTATGCGCTTTTGGCCCTCTCGGCTTGCTGCTTACCCGCTACAGGATG 192
OY	406 TTTTACATCTACAGAATTGCTGACCGCTTTACAGACCATCCTACCTCCAGAGTTGTAG 465
Dd	193 CCATTCATCCACCGCTTCGCGCAGCAAGCTTCTGGCGGCGCTCGGCCGACCTTCCTTG 252
OY	466 GACTCGTTGGCATCTTAAAGAAGCTGACATTTGCTGCAATTTTATCTACTCTGCAACT 525
Dd	253 GACTGTATGCTATCTCCTGAGATTTGGACCTCTGCTTATTCATATGTCATGGCGGAGCTAT 312
OY	526 GGGGAAGCGTATGCTATCTTAATGTTCTGGGATTTTGCTAATGAATAATTACAATAATCC 585

Db	313	GGGCGACGGCTGTTCCTCTCTCCTTCTTGSGGGTTTCCACACAGATCACACACTTG 372
QY	586	ACGAAGAACAAGCGTTTCTACGCCTTTTCGGTATCGAGCATATAATTCTTTACTAGCTT 645
Db	373	ATGACGCAAAAGATTCTACCTCTGTTCGGCCTTGGGGCAAAATATTGGCCCTTACTCTTT 432
QY	646	CTGGTCCTGTCACAATTGTTGGCGCTTAAGATGACAGCTCCCTTTCTGAAGGTGAGATC 705
Db	433	CCGGGCGTAGCTGTGAAGATATTTCTCAAATTTTCGGAGAACACATGGCGTCCAGGATTG 492
QY	706	CTTGGCAATTCTTTACCCTCTTTGATCGCAT 739
Db	493	GTTGGAGGATATCTTTGAAAGGAATGATGGCAT 526
RESULT	27	
LOCUS	B0147136	
DEFINITION	NF034F08FLP1074 Developing flower Medicago truncatula CDNA clone	
ACCESSION	B0147136	
VERSION	B0147136.1	
KEYWORDS	EST.	
SOURCE	barrel medic.	
ORGANISM	Medicago truncatula	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
AUTHORS	Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,	
TITLE	Flores,H.R., Imman,J.T., Weller,J.W. and May,G.D.	
JOURNAL	Expressed Sequence Tags from the Samuel Roberts Noble Foundation	
COMMENT	Medicago truncatula flower library Unpublished (2001) Contact: May GD Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel.: 580 221 7391 Fax: 580 221 7380 Email: gdmay@noble.org Insert Length: 700 Std Error: 0.00 Plate: 034 row: F column: 08 Seq primer: TCACACAGGAACACGATATGAC. Location/Qualifiers 1..700 /organism="Medicago truncatula" /db_xref="taxon:3880" /clone_id="NF034F08FL" /_flissue_type="Developing flowers" /_idev_stage="developmentally pooled. Contains a mixture of very young, developing, fully-opened flowers and flowers in early transition into pods." /_note="Vector: Lambda Zap; cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from very young, developing, fully-opened flowers and flowers transitioning into pods. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the GigaPack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using ExAssist helper phage and the E. coli strain XLI-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."	
FEATURES	source	
BASE COUNT	ORIGIN	
Query Match	9.1%:	Score 149.6; DB 14; Length 700;
Best Local Similarity	56.7%:	Pred. No. 1.7e-24;
Matches 331: Conservative	0:	Mismatches 248; Indels 5; Gaps 3;



```

QY 855 AACCTAAATGATATGAAAGTAGCTTCTCTATCTTGATAGATCTCTATATCTT 914
    ||||| ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 121 AAACCAAAATGGAACATGAGAGCTTGAATCTTGCTCTTCAAGTACATATTA 180
QY 915 TTATTACTCTCTGGTATATGCTTATGATATTTGCAATTAATCGAAGTACTGG 974
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 GATCAGACCACTTATGCTGCTTATGGAATCAGCATCAATCTTGATGAGTACATGG 240
QY 975 AAAATCGAGTGAACATATCTTATATGATGATATAGTACATGAGTCTCATGGGAAC 1034
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 AAGTCTAAGCTCAAGCTGATGCTTCCAGCCCCAACGAGTACTCTTCTTTCATGGGAGAC 300
QY 1035 TTCTCTCTGAGAGGCTAGTACTCCGACTTATCATGCTATTTGGTTGGTGTGTAACGTC 1094
    ||||| ||||| | | | | | | | | | | | | | | | | | | | | | |
Db 301 TTCTGACCGGCACTGG---AATTGCTATCATATATGATCTTCTAAGTCAATATATA 357
QY 1095 ATTCTAAATTTGGATGTTAAGTGGAGCCCTAGTCACTCTGTATGGTTCTCTTACA 1154
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 358 TTGACAATATGATGGAGGTGCTGCAAGATCACAGCTACAGTCTCTGCTTGAC 417
QY 1155 GGTATCGTTTCTTGCGCTTCTTATCTTTAGAAACCAAGCTTCTGGCTGGCTGCTATG 1214
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 418 GGAGTTGGTTCTCTCTCATATTTGGAGGTCCAAATTGCAACCCGCTCTTGATCA 477
QY 1215 TTGCGTACAACTCTCTCATGCTAGCTGTGTCGAGCTATACAGAAATATCTTGG 1274
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 478 ATTGGAATGACTCCACTGCTTACACTGATATATGTTGGCCATGACAGCAATTTTACG 537
QY 1275 AATTCACAAATATACGCTCTCTTACTCACTCACTAAAGAAATGCTTATCCCTCTTGGAC 1334
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 538 AAGAGTGTCTAAGTACAGTTTATTCACCCCTGC -AAGAAATGCTTACATTCCTTGGAT 596
QY 1335 CAAGGCAAAAAGTAAAGTAAAGCTGCTATGATATGATGATGCGCCGCTTGGAAAA 1394
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 597 GAATTAAGTAAAGTCAAGGGGAAAGCAGCATGATGTTGTTGACCC -AATGGGAAAA 655
QY 1395 TCAGAGGAGCTTATATCCAAAGATTTGCTGTTATCTGTG 1438
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 656 TCTGAGAGGAGCCCTTATCCACAGTTTCATGATCTTATTTGG 699

RESULT 28
LOCUS BF459810 499 bp mRNA linear EST 04-DEC-2000
DEFINITION 065F05 Mature tuber lambda ZAP Solanum tuberosum cDNA 5' similar to
plasmidic ATP/ADP-transp. . . g114138583, mRNA sequence.
ACCESSION BF459810
VERSION BF459810.1 GI:11528967
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 499)
AUTHORS Nielsen,K.L., Crookshanks,M., Emmersen,J. and Welinder,K.G.
TITLE EST-sequencing of mature potato tuber (Var. Kuras)
JOURNAL Unpublished (2000)
COMMENT Contact: Karen G. Welinder
Institute for bioteknologi
Aalborg Universitet
Sohnegaardhøjvej 49, 9000 Aalborg, Denmark
Tel: +45 96358467
Fax: +45 98141808
Email: kgw@bio.auc.dk
High quality sequence stop: 499
POLY-A-No.

FEATURES
SOURCE location/Qualifiers
1..499
/organism="Solanum tuberosum"
/cultivar="Field grown Kuras"
/db_xref="taxon:4113"

```

```

/clone_lib="Mature tuber lambda ZAP"
/tissue_type="tuber"
/Note="Vector: lambda ZAP"
BASE COUNT 132 a 105 c 114 g 148 t
ORIGIN
Query Match 9.0%; Score 148; DB 12; Length 499;
Best Local Similarity 58.1%; Pred. No. 4e-24;
Matches 281; Conservative 0; Mismatches 200; Indels 3; Gaps 1;

QY 934 TTGCTTATGCTATTTGATCATTAATCAATGAAGTCACTTGGAAAAAGTCAAGTCAAGCTGC 993
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 TAGCATATGGSCATTAGTATCAACCTTGTGAATATACAGTAAGTCAACCAAGCTCAAAAGCTC 60
QY 994 AATATCTTAATATGATGCTATAGTATGATGATGATGATGATGATGATGATGATGATGATG 1053
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 AGTTCCCAAGCCCCCAATGAGTACTCTCTATTCATGATGATGATGATGATGATGATGATGATG 120
QY 1054 TAGTATCCGCTACTTATCATGCTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1113
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 TAGGAACCTTTCACATGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 177
QY 1114 TAGTGGAGCCCTAGTCACTCTGTCATGTTCTCTCAACAGTATGCTTTCTTCTGCTC 1173
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 GAGCAGACCAAGATTAACACCTACAGTCTGCTCTTACCGGAGTTGGTTCTTCTTCTTCC 237
QY 1174 TTGTTATCTTTAGAAACCAAGCTTCTGGCTGTGCTATGTTGGTATGCTTCACTCTCTCA 1233
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 238 TCGTTTGGTTGGTGGACCTCAGACACCTACCTCTGCGAAGTTTGGAAATGACTCTCTTC 297
QY 1234 TCGTATGCTGTGGTTGCGGAGCTATACAGATATTTGTTGGAATCCACAAATATACGTC 1293
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 298 TAGCAGCTCTCTATGTTGGTGCATATGACAGACATTTTCTAGTATGAGTGCAGAGTATAGTT 357
QY 1294 TCTTTGACTCAACTAAGAAATGAGCTTATCTCTTGTACCAAGCAAGCAAAAAGTCAAG 1353
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 358 TGTTTGACCCCTGCAAGAAATAGGCTTACATCTTGTGATGAGACACCAAGGTTAAAG 417
QY 1354 GTAAAGCTCTCTTATGATGTTGCGCCGCTTGGAAAAATCAGAGAGAGCTTTATATCG 1413
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 418 GGAAGCAGCAATCGATGTTGTCGAATCACTGGGAAAGCTGAGAGCTTGTATAC 477
QY 1414 AACCA 1417
    || | |
Db 478 AACCA 481

RESULT 29
LOCUS B0507193 567 bp mRNA linear EST 22-JUL-2002
DEFINITION B0507193 Generation of a set of potato cDNA clones for microarray
analyses mixed potato tissues Solanum tuberosum cDNA clone STM6092
3' end; mRNA sequence.
ACCESSION B0507193
VERSION B0507193.2 GI:21923031
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 567)
AUTHORS Buell,C.R., Hartl,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Resterpe,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karameycheva,S.A.
TITLE Generation of a set of potato cDNA clones for microarray analyses
JOURNAL Unpublished (2002)
COMMENT On Jun 10, 2002 this sequence version replaced gi:21366062.
Other_ESTs: EST614607
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potatodetlgr.org

```



VERSION BM357837.1 GI:18426789  
 KEYWORDS EST.  
 SOURCE Triphysaria versicolor.  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Lamiales; Orobanchaceae; Triphysaria.  
 REFERENCE 1 (bases 1 to 529)  
 AUTHORS Matvienko, M., Torres, M.J. and Yoder, J.I.  
 TITLE Transcriptional responses to the hemiparasitic plant Triphysaria versicolor to host plant signals  
 JOURNAL Plant Physiol. 127 (1), 272-282 (2001)  
 MEDLINE 21437952  
 COMMENT Contact: Yoder, J.I.  
 John I. Yoder Research Lab, Dept. of Vegetable Crops  
 University of California at Davis  
 137 Asmundson Hall, One Shields Drive, Davis, CA 95616, USA  
 Tel: 530 752 1741  
 Fax: 530 752 9659  
 Email: jlyoder@ucdavis.edu  
 High quality sequence, stop: 530.  
 FEATURES  
 source  
 1..529  
 /organism="Triphysaria versicolor"  
 /db\_xref="taxon:64093"  
 /clone\_lib="Triphysaria versicolor root-tip, early  
 DMQ-induced transcript cDNA library"  
 /tissue\_type="root-tips -5mm length"  
 /dev\_stage="3-4 weeks growth"  
 /lab\_host="E. coli"  
 /note="Vector: PCR2.1 TA Cloning System, Invitrogen,  
 Carlsbad, CA; PCR-based suppression subtractive  
 hybridization cDNA library"  
 BASE COUNT 101 a 97 c 142 g 189 t  
 ORIGIN  
 Query Match 9.0%; Score 146.8; DB 13; Length 529;  
 Best Local Similarity 55.8%; Pred. No. 7.7e-24;  
 Matches 280; Conservative 0; Mismatches 222; Indels 0; Gaps 0;  
 294 TATGCAAGCTAAGTATATTTAAAGTAGAGAGCCCTATTATGAGGAGGAGCCG 353  
 Db TACACTAAGTGGCTAATGTTTGCAGAAAGATGCATTTCTACAGTATATGTTGCA 87  
 QY 354 TTTTAAATTTCTTGGCCCTGTCGCCGACTGAATTTATCCGCTACGCGATTTTACAT 413  
 Db TTTCAATGGCTTTTGGGGCATTTGGCTTATCCTCAGTAATTTTTCAC 147  
 QY 414 CCTACAGATTTGGTACCGCTTTACAGGCACTCCTACCTCCAGATTTGCTAGACTGGT 473  
 Db CCCACTGGGCTTGGTATAGCTTTTGAACGTGGGTCGAGGTTCTTGGACCACTT 207  
 QY 474 GCCATCTTAAAGACGTGACATTTGCTCATTTTACTTACTTGAACATAAGGGAGC 533  
 Db GCTATTTTGAAGATTTGAGGCTTTGTTGTCATGTTATGCGTGAGCACTTTGGGCGAGT 267  
 QY 534 GCTATGCTATGCTAATTTCTGGGGATTTGCTATGAATTTACAAAATCAGCAAGCA 593  
 Db GTGGTGTGTTCTGTTTGGGGGTTTGCAAATTCAGTAATCCACATGACGCAAAACC 327  
 QY 594 AAGCGTTTCTACGCTCTTTCGATATCGAGCTAATATTTTCTTTCAGCTTCTGTCGT 653  
 Db AAGAAATTCATCTCTTTTGGGCTCGGAGCAATGTAGACATCATTTCTCGGGTGG 387  
 QY 654 GCAATTTGTTGGGCTTCAAGTTGAGAGCTTCGTTTGAAGGTGATATCCTTTGGGGA 713  
 Db ACTGGAATTAATTTTCCCAATAGGCAAAATTTGGGCCCCGGTGTGATGTTGGGCC 447  
 QY 714 ATTCTTACGCTTTTATGAGCTATGACTATTTGATGCTGACTGTTCTATGCCCAGT 773  
 Db ATCTGTTGAAGGATATGATGTTTATGAGGGAGTGGCCCTATGATTTGTTTCTA 507  
 QY 774 TACTGTGATCAATTAAGACG 795

Db 508 TATGCTGGTAAACAATATATG 529  
 RESULT 32  
 C74932  
 LOCUS C74932  
 DEFINITION C74932 Rice cDNA from immature leaf including apical meristem Oryza sativa (japonica cultivar-group) cDNA clone E60959\_1A, mRNA sequence.  
 ACCESSION C74932  
 VERSION C74932  
 KEYWORDS C74932.1 GI:2443161  
 SOURCE EST.  
 ORGANISM Oryza sativa (japonica cultivar-group).  
 Oryza sativa (japonica cultivar-group).  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 469)  
 AUTHORS Sasaki, T. and Yamamoto, K.  
 TITLE Rice cDNA from immature leaf including apical meristem  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/PROJECT="RGP"  
 FEATURES  
 source  
 1..469  
 location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone\_lib="E60959\_1A"  
 /clone\_lib="Rice cDNA from immature leaf including apical meristem"  
 /dev\_stage="Immature"  
 /note="Organ: Leaf; Immature leaf including apical meristem (under long day condition)"  
 BASE COUNT 106 a 100 c 121 g 141 t 1 others  
 ORIGIN  
 Query Match 8.9%; Score 146.2; DB 14; Length 469;  
 Best Local Similarity 57.8%; Pred. No. 1.1e-23;  
 Matches 259; Conservative 0; Mismatches 189; Indels 0; Gaps 0;  
 1096 TTCTGAATTTGGATGTTAATGAGCCCTAGTCTCTGCTATGTTCTCTAAG 1155  
 Db TTAGGAAGTTTGGGTGGGAGTCGAGCTACATATACCTCTGCACTGTTGCTTCTCACTG 64  
 QY 1156 GTATGTTTTCCTCTCTTGTATCTTTAGAAACCAAGCTTGGCGTGGCTATGT 1215  
 Db GAGTTGGAATTCCTCTCACTGATTTTATTCGAGACCAATGATCTCTTATGAGGCACAG 124  
 QY 1216 TCGGTACACCTCTCTATGCTAGCTGTGTTGTGAGACTATACAGAAATTTCTTGA 1275  
 Db TTGGAATGACACTTTCCTTGGTGGGTCATATGTTGGGCAATGAGCAAACTTTTCAGTA 184  
 QY 1276 AATCCAAAATAGCCTCTCTTGTACATCACTAAGAAATAGGCGATATNCCCTCTTGACC 1335  
 Db AGAGTGCAGATGACAGTTTGTGACCTTGTAAAGAAATGSCATPACATTCCTTTGGACG 244  
 QY 1336 AAGACCAAAAAGTCAAAAGTCAAGCTGCTATTTGATGATGATGATGATGATGATGAT 1395  
 Db AGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 304  
 QY 1396 CAGAGGAGCTTTATCCAAACAGTTTGTCTGTTATCTGTGGAAGTATTTGAGCTATGA 1455  
 Db CTGGAGGGCGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 364  
 QY 1456 CCCCTATCTTCACTGATGATTTCTTTCATCATCATTTGATTTGTTCTGCAACTA 1515

```

Db 365 CACCTTACCTGGAGGAATTCATGTAATGCTTCTGCGTGGTGGCTGTAAGCT 424
OY 1516 AGTTAAACAACATATCTTAGGCGCAGTC 1543
Db 425 CCCTTGACTGCANAGTCTCCACCTGGC 452

RESULT 33
AM924705 581 bp mRNA linear EST 19-JUL-2000
LOCUS WS1_71.D05.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.
ACCESSION AM924705
VERSION AM924705.1 GI:8090530
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 581)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: water-stressed plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 528
POLYA-No.

FEATURES
source location/Qualifiers
1..581
/db_xref="taxon:4558"
/clone_id="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: Lambda zap; Site 1: XhoI;
Site 2: EcoRI; The library was made from poly-A RNA in the
cloning vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 149 a 125 c 147 g 160 t
ORIGIN
Query Match 8.9%: Score 146; DB 10; Length 581;
Best Local Similarity 58.9%: Pred. No. 1.2e-23;
Matches 251; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

OY 1118 TGAAGCCCTAGTCACTCTGTCATAGCTTCCTACAGAGATCGTTTCTTCGCTCTGT 1177
Db 11 TGCAGCTACATCAACACCTGAGTGTACTCTCACAGAGTGGGTTCTTCTCAGTCAT 70
OY 1178 TATCTTTAGAAACCAAGCTTTGGGCTGGTGGCTATGTTCCGTACAACTCCTTCATGCT 1237
Db 71 TTTGTTGGTAGCCATTCATGATACCAAGTTGGATGACACCTTGGCTTGC 130
OY 1238 AGCTGTGGTGTGGAGCTATACAGAAATATCTTTCGAAATCCAAATACGCTCTCT 1297
Db 131 GGCAGTCTATCTGAGAGCAATGCAAGACATTTTTCAGTAGAGTGCAAAAGTCAAGCTGTT 190
OY 1298 TGACTCACTAAGAAGATGGCTATATCCCTTGGACCAAGCAAAAAGTCANAAGTTAA 1357
Db 191 TGATCCCTGCAAGAGATGCGATACATTCCTTTGGATGAGATATGAAGGTAAAGTTAA 250
OY 1358 GCGTGTATGATGATGATGCTGGCGCCGCTTCGAAATACAGAGAGAGCTTAAATCAACA 1417

```

```

Db 251 AGCAGCTATATGACCTTGTCTCCACACCCCTGGGGAATCTGGAGGCTGTTGATCAGCA 310
OY 1418 AGTTTGCCTGCTTATCTGAGGAATGAGACTATAGACCCCTTATCTTTCGAGTATCT 1477
Db 311 GTTCATGATCTGTCATTTGATTTGATCTTCTTCCAGATTCGACCCCTTGGTGAATACT 370
OY 1478 TCTTTTCATTCATTTGCTATTTGGTTGTTCTTCACACTAGTTAAACAACATTTCTTACG 1537
Db 371 TCTGTTGATTTGTTCTTGTGATGCGTGGTGTGTAAGTGCCCTCAGACCTCAGATTTTCAAC 430
OY 1538 GCAGTC 1543
Db 431 CCTGCC 436

RESULT 34
BG322472 522 bp mRNA linear EST 27-FEB-2001
LOCUS EML_28.D05.b1_A002 Embryo 1 (EML) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.
ACCESSION BG322472
VERSION BG322472.1 GI:13152150
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 522)
AUTHORS Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 505
POLYA-No.

FEATURES
source location/Qualifiers
1..522
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_id="Embryo 1 (EML)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from lambda zap II; Site 1: XhoI; Site 2:
EcoRI; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 130 a 110 c 133 g 149 t
ORIGIN
Query Match 8.8%: Score 144.6; DB 12; Length 522;
Best Local Similarity 59.3%: Pred. No. 2.5e-23;
Matches 246; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

OY 1129 TCACCTCCGTCAAGTGTCTTCTACAGGATATGTTTCTTGGCTGTGTTATCTTGA 1188
Db 15 TCACCTCCGTCAAGTGTCTTCTACAGGATATGTTTCTTCTACATTTTGTGTC 74
OY 1189 ACCAAGCTTCTGGGCTGCTATGTTGCTACACTCTCTCATGCTAGCTGTGTTG 1248
Db 75 AGCATTGACCTCCTTATGACCAAGTTTGGATGACACCTTTCCTTGGCGCATTCATG 134
OY 1249 TCGAGCTATACAGAAATATTTTTCGAATTCACAAATACGCTCTTGTAGCTCAACTA 1308

```

Db 135 TTGGAGCAATGACAGACATTTTTCAGTAAAGATGCAAAAGTACGTCGTTTGATCCTTGA 194  
 Oy 1309 AAGAAATGACCTATATCCCTCTTGACCAAGAGCAAAAGTCAAGGTAAAGCTGATG 1368  
 Db 135 AAGAGATGGCATATCTCTTTGGATGAGATATGAGATTAAGAGTAAAGACCTATG 254  
 Oy 1369 ATGTAGTGGCCGCGCTGCGAAATGAGAGAGAGCTTTATTCACAAAGGTGCTCG 1428  
 Db 255 ACGTGTGCGCAACCCCTTGGGGAATGCGAGAGCTTGTGATCCAGCAGTTCATGTC 314  
 Oy 1429 TTATCTGGAAGATGAGAGCTATGACCCCTTATCTTGCACTGATCTCTTTTCATCA 1488  
 Db 315 TGTCAATTTGATCTCTTGGAATGCAACCTACCTTGGTGAATGATCTGAGATG 374  
 Oy 1489 TTGCAATTTGATCTCTTGGAATGCAACCTACCTTGGTGAATGATCTGAGATG 374  
 Db 375 TTCTTGCAATGCTGGGTGCTGTAAGGTCCCTGACCTCACAGTTTTCACACCTGGC 429

## RESULT 35

BF460035

LOCUS 072A12 Mature tuber lambda ZAP Solanum tuberosum cDNA 5' similar to  
 DEFINITION PLASTIDIC ATP/ADP-TRANSPORTER g1174417081p... sp1024381, mRNA  
 sequence.

ACCESSION

BF460035

VERSION BF460035.1 GI:11529192

KEYWORDS

SOURCE

ORGANISM

Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)  
 Contact: Karen G. Welinder  
 Institut for bioteknologi  
 Aalborg Universitet  
 Soingardsholmsvej 49, 9000 Aalborg, Denmark  
 Tel: +45 96358467  
 Fax: +45 98141808  
 Email: kwe@bio.auc.dk

High quality sequence stop: 744  
 POLYA-No.

Location/Qualifiers  
 1..744

FEATURES  
 source  
 /organism="Solanum tuberosum"  
 /cultivar="Field grown Kuras"  
 /db\_xref="taxon:4113"  
 /clone\_lib="Mature tuber lambda ZAP"  
 /tissue\_type="tuber"  
 /note="Vector: Lambda ZAP"

BASE COUNT 189 a 149 c 178 g 228 t  
 ORIGIN

Query Match 8.8%; Score 143.8; DB 12; Length 744;  
 Best Local Similarity 52.5%; Pred. No. 3.8e-23;  
 Matches 404; Conservative 0; Mismatches 337; Indels 28; Gaps 3;

Oy 571 AAATACAAAATCCAGCAAGACGTTTACGCTCTTTGCGTATCGAGCTAA 630  
 Db 1 AGATACACAGCTGCTGAGGCTAAGAGATTTCTATCTTTGTTGGACTTGGAGCGAATG 60  
 Oy 631 TTCTCTTACTAGCTTGTGCTGCTGCAATGTTGGGCTTCAAGTTGAGAGCTTCCGTTT 690  
 Db 61 TTGCTCTATTTTCTCTGCTGCGACAGTGAAGTACTTTTCTAGCTTGAAGACCTTTAG 120  
 Oy 631 CTGAAGGTGATCTCTTGGGGAATTTCTTACGCTTTTGTAGTGCCTATGACTATTGAT 750  
 Db 121 GTCTGAGATGATGTTGGGCAATCTCCCTGTAAGAGATGATGAGCATTTGTTGATGA 180

Oy 751 CTGACCTTGTCTTATGCGCCAGTACTGCTGATCAATAGAACGATATGACCATCTC 810  
 Db 181 TGGGTGGGCAATCTGTTCTTTTACTGGGTGATAGAAATGTTGCTTCCACTTC 240  
 Oy 811 GCTTCTATATATCCAGAGAAATGCAAAAGGGAAGAAAGGCTCTAAACCTAAATGAATA 870  
 Db 241 GTAGCAA-----GAGAGAGAAGGTAAACCTAACATGACCA 276  
 Oy 871 TCAAGATAGCTTCCCTCTATCTGATAGATCTCTTATATTTCTTTATTAACCTCTTGG 930  
 Db 277 CAATGAGACCTTGAAGTCTTGGTCTCTTCAAAATATATCAGGAGATCTTGGCACATGG 936  
 Oy 931 TTATGCGCATATGATTTGATTAATCAATAGAGTCTGGAAGTCTGGAAGTCTGGAAG 990  
 Db 337 TTGTAGCATATGACATTAATCAACCTTGTGAGATTAATCAATGAGTCAAGTCAAG 396  
 Oy 991 TGCATATCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1050  
 Db 397 CTCAGTCCCAAGCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 456  
 Oy 1051 GCGTAGTATCCGCTACTTATCATGCTATTTGTTGCTGATGATGATGATGATGATGAT 1110  
 Db 457 GAATGCAACTTTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 513  
 Oy 1111 GGTAACTGAGACCCCTAGCTACTGCTGATGATGATGATGATGATGATGATGATGATG 1170  
 Db 514 GGGAGACACACCAAGATATACACCTGATGATGATGATGATGATGATGATGATGATGATG 573  
 Oy 1171 CTCTGTTATCTTTAGAAACCAAGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1230  
 Db 574 CCTGCTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 633  
 Oy 1231 TCATGCTAGCTGTTGTTGCTGAGCTATGACAAATATTTCTTGAATCCACAAATACG 1290  
 Db 634 TTCTAGCAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 692  
 Oy 1291 CTCTGTTATCTTTAGAAACCAAGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1339  
 Db 693 GTTTGTTGACCCCTGCAAGAAATGGGCTACATCTTGTGATGAGA 741

## RESULT 36

BF053092

LOCUS EST438322 potato leaves and petioles Solanum tuberosum cDNA clone  
 DEFINITION CSTRB34M17 5' sequence, mRNA sequence.

ACCESSION

BF053092

VERSION BF053092.1 GI:10806988

KEYWORDS

SOURCE

ORGANISM

Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)  
 Contact: Cathy Renning  
 The Institute for Genomic Research  
 Division tel 1-800-711-6195, email cdna@esgen.com.

Location/Qualifiers  
 1..602

FEATURES

source

/organism="Solanum tuberosum"  
 /cultivar="Kennebec"  
 /db\_xref="taxon:4113"  
 /clone\_lib="potato leaves and petioles"  
 /tissue\_type="leaflets and petioles"  
 /dev\_stage="8 weeks old plants"



Db 400 GATCTTAACCTTTGGTCTACAGTAACTCACTCCGACTAGAGGGGCTCTTGCT 459

QY 1484 CATCATTCGTATTGGTGGTTCTTCGCACTAGTTAAACAACATATCTTACCGACATC 1543

Db 460 GATTCTTCTTGGCTGCTATATCAGCAACCAAGTCGCTGTATACAGATTACTGCTTGG 519

QY 1544 TGCCTTAAAGAACAGAGTGCCTCAGCAAGATTCAGTCTCTGCT 1589

Db 520 TCGAAGAGAGAACTTGAGATAGAGATGAGAGACGACGCTGCT 565

RESULT 38

LOCUS BF423978

DEFINITION BF423978 606 bp mRNA linear EST 06-DEC-2001

Accession BF423978

VERSION smg2f04.y1 Gm-cl047 glycine max cdna clone GENOME SYSTEMS CLONE ID:

KEYWORDS Gm-cl047-1784 5', similar to SW:TI01.SOLIU 024361 PLASTIDIC

SOURCE AT/ADP-TRANSPORTER.; mRNA sequence.

ORGANISM soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 606)

AUTHORS Shoemaker, R., Kelm, P., Vodkin, L., Erpelidg, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kueba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittner, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT Contact: Shoemaker R/public soybean EST project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccluresgen.com  
High quality sequence stop: 443.  
Location/Qualifiers  
1. 606  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl047-1784"  
/clone\_lib="Gm-cl047"  
/tissue\_type="Immature leaves (unfurled trifoliolate) of greenhouse grown plants"  
/dev\_stage="2 week old"  
/lab\_host="DH10B"  
/note="vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; This cDNA library was constructed from mRNA isolated from immature leaves (unfurled trifoliolate) of greenhouse grown plants that were 2 week old. The library was prepared using the Life Technologies psupertscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restriction site. SalI linker adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed in the laboratory of Dr. Lilia Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign (email l-vodkin@uiuc.edu)"

BASE COUNT 156 a 131 c 144 g 175 t

## ORIGIN

Query Match 8.7%; Score 143; DB 12; Length 606;  
Best Local Similarity 55.8%; Pred. No. 5.9e-23;  
Matches 294; Conservative 0; Mismatches 230; Indels 3; Gaps 1;

QY 975 AAAAGTCACCTCAAGCTGATATCTTATATGATAGATATAGTATGATGAGGGAAC 1034

Db 1 AATATCAAGCTCAAGAGCTAGTTTCTTACGCCCAAGATGACTCATCTTTATGGAGAC 60

QY 1035 TTCTCCTTCTGAGCGGCTAGTATCCGACTTATATCATATTTGGTGGTAAAGTC 1094

Db 61 TTCTCAACTGCAACAGGATGCTACATTCATTAATGATGTC--TTCTAGTCAATTATA 117

QY 1095 ATTCTGAATTTGATGTGTTAACTGAGCCCTAGTCATCTCTGTATGTTCTCTTACA 1154

Db 118 TTGGCAAAATATGATGGAGGGAGTGGTGGCAAGATCACCCCAAGCTCTCTGATATGACA 177

QY 1155 GGTATCGTTTCTCTGCTCTGTTGTTATCTTACGAACCAAGCTTGGCGCTGCTATG 1214

Db 178 GGAGTTGGTTCTTTCTTCTTATATATTGCGCGCCCAATGCTCTTATTTGCAAG 237

QY 1215 TTGCGTACAACTCTCTCATGCTAGCTGCTGTTGCGAGCTATACAGATATTTCTG 1274

Db 238 TTGGATGATGCTCCACTGCTAGCTGCTGTATATGGGCGCATGCAAGAAATATTCAGC 297

QY 1275 AATTCACAAATATGCGCTCTTTGACTGACATAAAGAAATGAGCTATTCCTCTTGAC 1334

Db 298 AAGAGTGTAACTAGTACAGTTTATTCGATCCGTCGCAAGAAATGCGCCATACCTTAGAT 357

QY 1335 CAAGAGCAAAAGTCAAGAGTATGAGTGTATGATGTTGGCGCGCTTGGGAAA 1394

Db 358 GAGGATACCAAGGTCAAGGAAAGGACCATGATGATGTGTGTCAGCAACCATGAGAAA 417

QY 1395 TCAGAGAGAGCTTATATCCACAAAGTGTCTGTTATCTGTGAATATGAGCTATG 1454

Db 418 TCTGGAGAGAGCTGTGATTCAGAGCTCATGATCTTATGTTGTTGTTGTTGTTGTTGTT 477

QY 1455 ACCCTTATCTTGCACTGATTTCTTTTTCATCATTTGCTATTTGGTT 1501

Db 478 GCCCATACCTAGAGAGGGGCTCTGTGATGTTCTTGGCGGCT 524

RESULT 39

LOCUS BG456211 663 bp mRNA linear EST 19-MAR-2001

DEFINITION BG456211 NF075B07P.L1059 Phosphate starved leaf Medicago truncatula cDNA

ACCESSION BG456211

VERSION BG456211.1 GI:13379536

KEYWORDS EST.

ORGANISM barrel medic.

SOURCE Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.

REFERENCE 1 (bases 1 to 663)

AUTHORS Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.

TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula phosphate-starved leaf library

JOURNAL Unpublished (2000)

CONTACT: Harrison MJ

Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7325  
Fax: 580 221 7380  
Email: mjharrison@noble.org  
Insert Length: 663 Std Error: 0.00  
Plate: 075 Row: B Column: 07  
Seq primer: TCACACAGAAACACCTATGAC.



FEATURES	Source	Location/Qualifiers
		1..663
		/organism="Medicago truncatula"
		/db_xref="taxon:3880"
		/clone="NF075807.P1"
		/clone_lib="phosphate starved leaf"
		/tissue_type="leaf"
		/dev_stage="trifoliolate"
		/note="Vector: Lambda Zap: At the trifoliolate stage, M. truncatula plants were transplanted to phosphate-free sand and grown for a further 30 days. During this 30 day period, the plants were fertilized twice weekly with 1/2 Hoaglands solution containing only 200m potassium phosphate. RNA was prepared from above ground tissues."
BASE COUNT	181 a	140 c 156 g 186 t
ORIGIN		
Query Match	8.7%	Score 142.8; DB 12; Length 663;
Best Local Similarity	54.9%	Pred. No. 6.6e-23;
Matches	282; Conservative	0; Mismatches 232; Indels 0; Gaps 0;
QY	1074	CAATTGTTGGTGGTGAAGCTCATTCGTAAATTTGGATGGTGTACTGAGAGCCCTAGTCACT 1133
DB	4	CGAGTTTCTAAGTCAATATATATATTACAAATATGAGTGGGAGTTGCTGCCAAGATCACA 63
QY	1134	CTGTCATGTTGGTTCCTCTCTTAACAGATATCGTTTTCCTGCGCTTGTATCTTTATGAAACAA 1193
DB	64	CCTACAGTCCTCTCTCTGACCGGAGTTGGTTTCTCTCTCATATGTGTTGGAGTCCA 123
QY	1194	GCTTCTGGGCTGGTGGCTATGTTGGGTACACTCTCTCATGCTATGCTGTGGTTGGGA 1253
DB	124	ATTGCAACCCGGCTTGCATCAATTTGGAATGACTCCACCTGCAAGCAGCTGATATCTTGCT 183
QY	1254	GCTATACAGAAATATTTCTTTGCAATTCACAAATATCGCTCTGTCAGTCAACTTAAAGA 1313
DB	184	GCCATGACGAACATTTTATAGCAAGAGTGTACTAGTATTTATTTGACCCCTGCACAAAGA 243
QY	1314	ATGGCCTATATCCCTCTTGACCAAGAGCAAAAGTCAAAAGGTAAAGCTGATATGATGTA 1373
DB	244	ATGGCTTACATTTCCCTTGGATGATAGTACTAGTCAAGGCGGAAGCAAGCATTTGATGTT 303
QY	1374	GTTGCGCCGCCCTTGGGAAATCAGAGAGCTTTAATCCAAACAGTTTGTGCTGTTATC 1433
DB	304	GTTTCAACACCCATTTGGGAAATCTGGAGAGCCCTTATCAACAGTTCATGATCTTAAT 363
QY	1434	TGTGAAGATTTGGAGCTATGACCCCTTATCTTGGAGATGATCTTCTTTCATCTAGTCT 1493
DB	364	TTTGGTTACTATGCAACTCACTCACTCATACCTTGGAGGTGTGCTTCTTGTGATGTTCTT 423
QY	1494	ATTGTTGGTTGTTTTCGCAACTAAGTTAAACAACATATCTTAAAGCCAGCTGCTCTTAAA 1553
DB	424	GCTCTGTTTAGAGGACGACCAAGTCTTTGGATACACAGTTTACTGCAATTCGCTCAGAGGAA 483
QY	1554	GACCAAGAAGTGGCTCAAGAAGATTCAGCTCTG 1587
DB	484	GAACTCGAGAAAAACAGATGAAAAAGACACAGCAGCTG 517
RESULT 40		
BE726529		
LOCUS	BE726529	575 bp mRNA linear EST 14-SEP-2000
DEFINITION	894092JF04.y2 C. reinhardtii CC-1690, normalized, lambda Zap II	
ACCESSION	BE726529	
VERSION	BE726529.1	GI:10127825
KEYWORDS	EST.	
SOURCE	Chlamydomonas reinhardtii.	
ORGANISM	Chlamydomonas reinhardtii.	
REFERENCE	1 (bases 1 to 575)	
AUTHORS	Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., Mcdermott,J.P., Sillflow,C., Stern,D. and Surzycki,R.	

TITLE		ANALYSES OF THE CHLAMYDOMONAS REINHARDTII GENOME: A MODEL, UNICELLULAR SYSTEM FOR ANALYZING GENE FUNCTION AND REGULATION IN VASCULAR PLANTS; PROJECT PHASE 2	
JOURNAL		UNPUBLISHED (2000)	
COMMENT		CONTACT: CHARLES HAUSER DCMB BOX 91000 DUKE UNIVERSITY DURHAM, NC 27708-1000 TEL: 919 613 8159 FAX: 919 613 8177 EMAIL: CHAUSER@DUKE.EDU	
FEATURES		LOCATION/QUALIFIERS	
SOURCE		1..575	
		/organism="Chlamydomonas reinhardtii"	
		/strain="CC-1690 Wild type mt+ 21gr"	
		/db.xref="taxon:3055"	
		/clone.lib="C. reinhardtii CC-1690, normalized, lambda Zap II"	
		/note="Vector: Bluescript II SK-; Site.1: EcoRI; Site.2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldi et al (1996) Genome Research 6: 791-806."	
BASE COUNT		103 a 201 c 158 g 113 t	
ORIGIN			
Query Match	8.6%	Score 141:	DB 12: Length 575:
Best Local Similarity	54.1%	Pred. No. 1.7e-22:	
Matches 310: Conservative	0:	Mismatches 260:	Indels 3; Gaps 1:
855	AAACCTAAATAATATGAAGAATAGCTCCCTATCTGTAGATGATTCCTATATTCCT	914	
DB	5 AAGACCMAGATBAGCGTGGGCGAGTCTTCGCTTCGCGCACTCGAGCTACATCCGC	64	
915	TTATTTAACTCTCTGTTATTTGCTATGTTGATTTGCAATTAATGAAAGACTGTTG	974	
DB	65 GACATGGCCACCTGATGGTGCCTACGAGTATCTGCATCACTGATGAGCTGAGTACCTGG	124	
975	AAAAGCTACGCTAAACTGCAATATCTCTAATATGATGACTATATGAGAGTTCATGGGGAAC	1034	
DB	125 AAGGGCAAGCTCAAGGCCCAAGTTCCCAACCCCAAGCAAGCACTCTGCTGTTGGGCGAG	184	
1035	TTCTCTCTTGAGCTGGGGGTAGTATTCGCTATTCATGATGCTATTTGTTGGTGTGAACGTC	1094	
DB	185 TTCTCCACCGCCACCGTACCTGACCTTACCAATGATGATCTCTTCCCGCTGGATCTTC	244	
1095	ATTCTAAATTTGATGTGTTAACTGGAGCCCTAGTCACTCTGTCATGATGTTCTCTTACA	1154	
DB	245 AAGC---AGTTGGCTGGGGGTGTGGCTGCCGTGATACACCCCATCATGCTGTGATCAC	301	
915	GGTATTCGTTTCTTGCTGCTTCTGTTATTCCTTTAGAAACAAGCTTCTGGGCTGCTGATG	1214	
DB	302 GGTCTGCTGTTCTTACCTTGCTGCTGGCGGCGAGACCGCTCAAGCCCGCTTGCGCCGC	361	
1215	TTTCGGTACACGCTCTCATGCTTACTCTGGTTGGTGGAGGCTATACAGATATCTTTTCG	1274	
DB	362 TGGGCGATGACCCCTCTTACGCGCCGCGTGGTGGGGGCGGCCAAGACGCTCTTCTCC	421	
1275	AAATCACAATAATAGCTCTCTTTTACTCAACTAAGAANAAGGCTATATCCCTTTGAC	1334	
DB	422 AAGTGTCTCAAGTACTCCTCTGTTCAGACCCCTGCAAGAGATGCGCTACATCCCCCTGGAG	481	
1335	CAAGAGCAAAAAGTCAAGATGAGGCTGCTATTGATGTAGTTGGCCGCCCTTGGAAAA	1394	
DB	482 GACGAGGTCAAGACCAAGGCGACCGCCATATGACGTGATTCGACACCGCTGGGCAAG	541	



Wed Jan 22 10:21:05 2003

us-09-869-433-1.rst

Page 27

Oy	1395	TCAGGAGGAGCTTTAATCCACAAGGTTTCTC	1427
Db	542	TCCGGGCGTGCCCTGATCCAGCAGTTCATGATC	574

```
Search completed: January 21, 2003, 19:27:17
Job time : 1706 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 19:36:33 : Search time 2952 Seconds  
(without alignments)  
16138.649 Million cell updates/sec

Title: US-09-869-433-1

Perfect score: 1637

Sequence: 1 gaataaanaactacagaa.....gttgatccctactgctttt 1637

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl:\*

1: gb.pa:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.or:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*  
29: em.vi:\*  
30: em.htg.hum:\*  
31: em.htg.inv:\*  
32: em.htg.other:\*  
33: em.htg.mus:\*  
34: em.htg.pln:\*  
35: em.htg.rod:\*  
36: em.htg.man:\*  
37: em.htg.vtc:\*  
38: em.sy:\*  
39: em.htgo.hum:\*  
40: em.htgo.mus:\*  
41: em.htgo.other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	1535	93.8	13389	1	AE002202	AE002202 Chlamydia
2	1484	90.7	11421	1	AE001619	AE001619 Chlamydia
3	1484	90.7	300650	1	AP002546	AP002546 Chlamydia
C 4	45	2.7	23773	1	AE002301	AE002301 Chlamydia
5	39	2.4	1587	1	TAJ10586	TAJ10586 Chlamydia
C 6	39	2.4	14168	1	AE001281	AE001281 Chlamydia
7	21	1.3	5032	6	AX281684	AX281684 Sequence
8	21	1.3	8220	6	AR064709	AR064709 Sequence
9	21	1.3	8220	6	AR089156	AR089156 Sequence
C 10	21	1.3	17462	6	AL592487	AL592487 Human DNA
11	21	1.3	19124	3	PFPAVARIA	PFPAVARIA
12	21	1.3	19124	6	AR089157	AR089157 Sequence
C 13	21	1.3	90106	2	AC007499	AC007499 Homo sapi
14	21	1.3	98286	2	AC092392	AC092392 Homo sapi
C 15	21	1.3	126996	2	AC113803	AC113803 Rattus no
16	21	1.3	134482	9	AC007335	AC007335 Homo sapi
C 17	21	1.3	143279	9	AC104838	AC104838 Homo sapi
18	21	1.3	150495	2	AC046148	AC046148 Mus muscu
C 19	21	1.3	150720	2	AL513016	AL513016 Human DNA
20	21	1.3	138599	2	AC101941	AC101941 Mus muscu
C 21	21	1.3	165679	2	AC128538	AC128538 Homo sapi
22	21	1.3	171747	9	AP001554	AP001554 Homo sapi
C 23	21	1.3	173595	2	AC020766	AC020766 Homo sapi
24	21	1.3	206463	2	AP000834	AP000834 Homo sapi
25	20	1.2	1770	6	AX098422	AX098422 Sequence
26	20	1.2	2110	8	AF428316	AF428316 Arabidops
C 27	20	1.2	2151	8	AY045903	AY045903 Arabidops
28	20	1.2	2181	8	ATANTMR	249227 A. thaliana
C 29	20	1.2	2762	2	AF222985	AF222985 Homo sapi
30	20	1.2	7291	9	AF222983	AF222983 Homo sapi
C 31	20	1.2	15002	9	AF222981	AF222981 Homo sapi
32	20	1.2	33376	9	AF222987	AF222987 Homo sapi
C 33	20	1.2	33780	6	AX151213	AX151213 Sequence
34	20	1.2	48546	9	AC107059	AC107059 Homo sapi
C 35	20	1.2	57070	2	AL354721	AL354721 Human DNA
36	20	1.2	58810	2	AC113541	AC113541 Mus muscu
C 37	20	1.2	65259	2	AC113102	AC113102 Mus muscu
38	20	1.2	74010	2	AC114147	AC114147 Rattus no
C 39	20	1.2	74274	2	AC087717	AC087717 Homo sapi
40	20	1.2	77422	3	AC025723	AC025723 Caenorhab
C 41	20	1.2	86897	9	AC093843	AC093843 Homo sapi
42	20	1.2	95537	2	AC021512	AC021512 Homo sapi
C 43	20	1.2	103984	2	AC129468	AC129468 Rattus no
44	20	1.2	104289	2	HS104D14	HS104D14
C 45	20	1.2	111963	2	AC099180	AC099180 Rattus no
46	20	1.2	120387	10	AL604046	AL604046 Mouse DNA
C 47	20	1.2	121637	8	AC122544	AC122544 Genomic s
48	20	1.2	122897	2	AP001953	AP001953 Homo sapi
C 49	20	1.2	127491	2	AC124962	AC124962 Medicago
50	20	1.2	132990	8	AC018848	AC018848 Arabidops
C 51	20	1.2	133968	9	AL116171	AL116171 Human DNA
52	20	1.2	136284	9	HS53C18	HS53C18
C 53	20	1.2	139505	9	HSJ591C20	HSJ591C20 Human DNA
54	20	1.2	145052	10	AL603889	AL603889 Mouse DNA
C 55	20	1.2	145616	2	HS108K11	HS108K11
56	20	1.2	147517	2	AC102030	AC102030 Mus muscu
C 57	20	1.2	150651	2	AC016535	AC016535 Homo sapi
58	20	1.2	150681	2	AC006011	AC006011 Homo sapi
C 59	20	1.2	150902	2	AC098652	AC098652 Homo sapi
60	20	1.2	152118	2	AC013263	AC013263 Homo sapi
C 61	20	1.2	156255	2	AP002369	AP002369 Homo sapi
62	20	1.2	157527	9	AC117746	AC117746 Homo sapi
C 63	20	1.2	158524	2	AL162372	AL162372 Homo sapi
64	20	1.2	159991	9	AC093628	AC093628 Homo sapi
C 65	20	1.2	165739	9	AL353728	AL353728 Human DNA

C 66	20	1.2 165951	9	AC010730	AC010730 Homo sapi
C 67	20	1.2 166308	9	CNS01DX2	AL139020 Human chr
C 68	20	1.2 171782	2	AC095121	AC095121 Rattus no
C 69	20	1.2 171847	9	AL590229	AL590229 Human DNA
C 70	20	1.2 176044	2	AC073845	AC073845 Homo sapi
C 71	20	1.2 179391	2	AC115195	AC115195 Rattus no
C 72	20	1.2 179484	2	AC098070	AC098070 Rattus no
C 73	20	1.2 180210	2	AC109156	AC109156 Mus muscu
C 74	20	1.2 182653	2	AC130220	AC130220 Mus muscu
C 75	20	1.2 182915	2	AC019213	AC019213 Homo sapi
C 76	20	1.2 186493	2	AC125550	AC125550 Rattus no
C 77	20	1.2 187461	9	AL590453	AL590453 Human DNA
C 78	20	1.2 187490	9	AC023902	AC023902 Homo sapi
C 79	20	1.2 187638	2	AC118273	AC118273 Homo sapi
C 80	20	1.2 196514	9	CNS05TBP	AL163952 Human chr
C 81	20	1.2 197592	2	AC026232	AC026232 Mus muscu
C 82	20	1.2 198280	9	AC092506	AC092506 Homo sapi
C 83	20	1.2 198784	10	AC091521	AC091521 Mus muscu
C 84	20	1.2 199514	10	AC068902	AC068902 Mus muscu
C 85	20	1.2 205284	2	AL732520	AL732520 Mus muscu
C 86	20	1.2 209436	9	AC026227	AC026227 Homo sapi
C 87	20	1.2 210516	2	AC091274	AC091274 Mus muscu
C 88	20	1.2 219054	2	AC068663	AC068663 Mus muscu
C 89	20	1.2 248865	2	AC074003	AC074003 Homo sapi
C 90	20	1.2 298216	2	AC006875	AC006875 Caenorhab
C 91	20	1.2 346940	2	AC107420	AC107420 Homo sapi
C 92	19	1.2 155	11	G64013	G64013 2156SP6 Hum
C 93	19	1.2 626	3	AF340093	AF340093 Biorhiza
C 94	19	1.2 626	3	AF340096	AF340096 Biorhiza
C 95	19	1.2 840	1	AHLEXA	X77263 A.hydrphil
C 96	19	1.2 985	11	CNS06KOP	AL403151 77 end of
C 97	19	1.2 1001	6	AX081559	AX081559 Sequence
C 98	19	1.2 1101	6	AX374740	AX374740 Sequence
C 99	19	1.2 1185	5	VUTMTDG	L33397 Vultur gryp
C 100	19	1.2 1186	5	GYWMTDG	L33379 Gymnogyphs c

## ALIGNMENTS

RESULT 1  
AE002202/c 13389 bp DNA linear BCT 30-MAY-2000  
LOCUS Chlamydomphila pneumoniae AR39, section 33 of 94 of the complete  
DEFINITION

ACCESSION AE002202 AE002161  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS Chlamydomphila pneumoniae AR39.  
Chlamydomphila pneumoniae AR39

REFERENCE  
AUTHORS Bacteria: Chlamydiales: Chlamydiaceae: Chlamydomphila.  
1 (bases 1 to 13389)

REFERENCE  
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,J.G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
pneumoniae AR39

JOURNAL  
MEDLINE Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
PUBMED 20150255  
10684935

REFERENCE  
AUTHORS 2 (bases 1 to 13389)

JOURNAL  
MEDLINE Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,J.G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Direct Submission

JOURNAL  
COMMENT Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
On Jun 1, 2000 this sequence version replaced gi:7189324.  
FEATURES Location/Qualifiers

source	1. .13389 /organism="Chlamydomphila pneumoniae AR39" /strain="AR39" /db_xref="taxon:115711" /note="synonym: Chlamydia pneumoniae AR39" 163. .339 /gene="CP0400" 163. .339 /gene="CP0400" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AF38245.1" /db_xref="GI:7189325" /translation="MLSVSKLLIPDSDIERETDSRRDKKIFITYICSSKVLAGEFFS HLDRHKRIHSTIGV" 597. .1448 /gene="CP0401" 597. .1448 /gene="CP0401" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AF38246.1" /db_xref="GI:7189326" /translation="MVNIQPVYRNQVNSQATQFSCVPALSLITVSVAALVIA LVCSQSLSELTALVLSLIFASAMFIMKROPEKELIPKIMELIDENPSI VDFPIRQEVSIYEIHHLLISLTKTKVFPKAPYLDKELQGIKFKVHDSKLPNI EETLHCHPMLHGLRIVPMSVPTGTGYWCGFLGYENAPSLFEERRSLILKKI SRGEFALIEDGLKKNWSSSELVQIQNLPTRYVADKEVDENLADYQORDSLHL IRSHKLS" 1455. .1808 /gene="CP0402" 1455. .1808 /gene="CP0402" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AF38247.1" /db_xref="GI:7189327" /translation="MQLROYMNSGMDWICDDEDSQSGEGFQLSRTVCLLSSWATLEAK EORFLPEVSLTWEELEIMQLSKPTKHGAKDLGVFERHFORQYIGSLDNLORF ENFELNYPKYHLDRE" 1842. .1958 /gene="CP0403" 1842. .1958 /gene="CP0403" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AF38248.1" /db_xref="GI:7189328" /translation="MKSLSVRSFAFSGIVKSLPNLKGKRELFPFGFTTNGS" 2019. .3326 /gene="CP0404" 2019. .3326 /gene="CP0404" /note="hypothetical protein; identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AF38249.1" /db_xref="GI:7189329" /translation="MAMSWTIVPHALFKNHKCEHSTFPLSRITVIRIAIASLFCIGAL
--------	---

AAACCLAPVSVYGVSLAFIAVLLSLVIALIFEGEKKLPPTPRRIIPDRFTHVIDEA  
YGLSISAFVREOQVTLAEFRQSTFALCNISPEEKIKOLPELSKSVESFISRLAGD  
LEKNMNPJLEEDLSOTCPYLWLOKIFISAGDPOCVLEPRECGYVWGLSGSTAK  
ATICKETHHTILOOTLKEDVLLKNNKAQEMWDDEWKAIVERITYTTTAGTILKTEA  
GGIKRETI SKELLELSLSTGYSPDOLQTLQPRADMWLCEVDNSTAINODCALVGA  
LSSONLDESSIDFVDNGLYVLDLKEAVOAFSADBPKEKELGFLLRHDSVSKRL  
BSYRQGLHRIALDHGNARARYDVNFVTGARIRHKSIFPKD"  
/complement(3404. .4795)  
/gene="CP0405"  
/complement(3404. .4795)  
/gene="CP0405"  
/complement(3404. .4795)  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAF38250.1"  
/db\_xref="GI:7189330"  
/translation="MLNLRYMTQIDPRKMSNTSPYTONNRSCNYFEELKNTTII  
IVISAILLGLALIAFLCYAAPVSYLSGALLGLLIGVIGIKITPMISKEO  
VPEQVLRIRAHYPKFVSDFVEAKPKLKDLSIFDILLNOLHVSSTSTNSSEL  
OOKIDFEGIRAKLKNVPTASLKRLESASRPPLPSLAKILQKYFPFWMGCEPISAG  
SKVVELHRVKTIGSLDESDSYTKPEMLPTWLIPLDPRPTNSILNLHTLVLAARL  
TRDVFQHLKYALANGEMNLHSDLNTMKQOLFARYHAAYQSYKHLQSPLOEDXYNL  
LDCIFKRYSWKMSLITKTVPADLMENLCLDHTGRDOMEFASLIGTLYOGLIH  
KSEAFSLSLFLSIDOKFTIRROSTINAMLEMLATHNSTFRSLPPTTHPLKRSVF  
SQPEDSSLLIG"  
complement(4776. .5021)  
/gene="CP0406"  
complement(4776. .5021)  
/gene="CP0406"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAF38251.1"  
/db\_xref="GI:7189331"  
/translation="MRFRNKSLIFIKIRIYSOSGKEOKARPPFKKSTISLVILL  
LEAIFENFSSITONFNKRNKNKINISINRIFVFTT"  
complement(5081. .6361)  
/gene="CP0407"  
complement(5081. .6361)  
/gene="CP0407"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAF73665.1"  
/db\_xref="GI:8163426"  
/translation="MRHSGCTMKCSPLTLVHIFLKNDCBCHRSCLKIRIARILLIG  
LVNLVLSLSEVFLAAPTISVAIGSTLAAAVLIIITLVALLKSKVLPIPNELQKI  
LYNRYPKVEFVETHTSLTVNELKIFINCKMSGDPLPNLHKKEAFGLKSIDTL  
LFPEFEELILQNCPLWYLSHFIDKTESVAGSIGLNKQKVGGLAPLHKGYYTIIH  
SYTRPLTLISESOYKFLYSKASKNOMQSPVKTCCEITFELPHNMIFRQDVQGISO  
FLLFESGTMBOAOMTOLINPDNKKMLCOFDKARGCSNATGCGPLNTEPNMDEY  
SSNVEPYNFTWOBELKYLEKESPMHPSALVOKICVNTTHONLKRMOEVRNT  
SSQWTSLLPQYAFHAQYKLEKIESSLPRSSL"  
complement(6479. .8026)  
/gene="CP0408"  
complement(6479. .8026)  
/gene="CP0408"  
/note="similar to GB:M28816 SP:P19568 PID:152470  
GB:AJ23269; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="ADP, ATP carrier protein"  
/protein\_id="AAF38252.1"  
/db\_xref="GI:7189332"  
/translation="MTKTEKRPCKLSLFWPIHTHELKVLPMELMFCITFTNYTL  
RDRDITLVGAPSGAALIPKIFKVLVWPCAIIFMLYAKLSNLSKALFEYAAGTPE  
LIPALFPTVIYLRDVLHPTEFADRLQALIPGLLGVALLRMWTFAAFVLAELMG

SVMLSMFGEFANETTKIHEAKREYALFGICANISILASGRATVWASKLRASVSGVD  
PMSILRLMAMTIVSGYLWASWIMKNNLTDREYNPEMKGGKGA PKNMKMD  
SPVLARSPLYLILALVYATGICNMLEVWYKSLKIQYPRMNDYSEFKNFSWTG  
VSVYILMFVGANVIRKFGWLGALVTEVWVLGTYEFAVIFRNQASGVAMGTT  
PLMLAVVGAIONISKSTKATLADPSTEMAYTLPIDODOKYKGAIDIVARFGSG  
GALTOOGLIGSIGAMTPLYAVILFIIAIWVSATRLNKLFLAQALKEOVAOE  
DSAPMS"  
complement(8227. .8742)  
/gene="CP0409"  
complement(8227. .8742)  
/gene="CP0409"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAF38253.1"  
/db\_xref="GI:7189333"  
/translation="MGEVAVROSHIKKEIEKLIGKAIKRVCGKNKENDLCRYLPSPSG  
YMHFTTKKMSAPPEOLKMLKTFIESEPRTINPRPARPGSKRRRDIINFTKT  
IERVLELARQYGDLDLARFSPKRPRLTSLKRELLIRISIRNGIVSELMAVYEAARAVS  
SPNLEVISPEV"  
complement(8910. .9113)  
Query Match 93.8%; Score 1535; DB 1; Length 1389;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1635; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
gene  
1 GAAATAAAAAACATACAGATAGAAAATAAAGATATTCAGAGGTAATATACAAAAA 60  
|||||  
8076 GAAATAAAAAACATACAGATAGAAAATAAAGATATTCAGAGGTAATATACAAAAA 8017  
|||||  
QY 61 CCGAAGAAAAACCTTTGGAAAAATTCGCTCTTCTTGGCCGATACATCTACGAGC 120  
|||||  
Db 8016 CCGAAGAAAAACCTTTGGAAAAATTCGCTCTTCTTGGCCGATACATCTACGAGC 7957  
|||||  
QY 121 TAAAGAAAGTTGCGCAATGTCCTAATGTCCTGATTCATTTACTATAGGTTG 180  
|||||  
Db 7956 TAAAGAAAGTTGCGCAATGTCCTAATGTCCTGATTCATTTACTATAGGTTG 7897  
|||||  
QY 181 TACGGGATACAAAGACACTTATGTCGAGACTCTGCTCTGTCGAGAGGCAATAC 240  
|||||  
Db 7896 TACGGGATACAAAGACACTTATGTCGAGACTCTGCTCTGTCGAGAGGCAATAC 7837  
|||||  
QY 241 CTTTCATCAAGTTTGGCTGTGTCCTGTCCTGTCATTAATCTTATATATGCA 300  
|||||  
Db 7836 CTTTCATCAAGTTTGGCTGTGTCCTGTCCTGTCATTAATCTTATATATGCA 7777  
|||||  
QY 301 ACCTAAGTAATTTTAAGTAGCAGGCGCTTATTTATGCACTGGGAACGCCCTTTTAA 360  
|||||  
Db 7776 ACCTAAGTAATTTTAAGTAGCAGGCGCTTATTTATGCACTGGGAACGCCCTTTTAA 7717  
|||||  
QY 361 TTTTCTTGGCCCTGTTCCCGACTGAATTTATCCGCTACGAGTGTTTTACATCTACAG 420  
|||||  
Db 7716 TTTTCTTGGCCCTGTTCCCGACTGAATTTATCCGCTACGAGTGTTTTACATCTACAG 7657  
|||||  
QY 421 AATTTGCTGACGCTTTTACAGGCGCATCTACCTCAGAGATGCTAGACCTGTTGCCATCT 480  
|||||  
Db 7656 AATTTGCTGACGCTTTTACAGGCGCATCTACCTCAGAGATGCTAGACCTGTTGCCATCT 7597  
|||||  
QY 481 TAGAAGACGAGCATTTGCTGATTTTATGCTAGTGAAGTATGAGGGAACGCTATGCG 540  
|||||  
Db 7596 TAGAAGACGAGCATTTGCTGATTTTATGCTAGTGAAGTATGAGGGAACGCTATGCG 7537  
|||||  
QY 541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAAATCCAGAGCAAGCGTT 600  
|||||  
Db 7536 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAAATCCAGAGCAAGCGTT 7477  
|||||  
QY 601 TTTACGCTCTTTTGGGATTCGAGGCTAATATTTCTTTACTAGCTTCTGCTGCAATTG 660  
|||||  
Db 7476 TTTACGCTCTTTTGGGATTCGAGGCTAATATTTCTTTACTAGCTTCTGCTGCAATTG 7417  
|||||  
QY 661 TTTGGGCTTCAAGATTGAGAGCTTCCGTTTCTGAAGGATGATGATCTTGGGAAATTTCTT 720

```

|||||
Db 7416 TTTGGGCTTCAAGTTGAGAGCTTCGTTCTGAAAGCTAGATCCCTGGGAAATTTCTT 7357
Oy 721 TACGCTTTTGGTGGCTATGATGTATGCTGGAAGCTTCTCTTATGAGCACTTACTGCT 780
Db 7356 TACGCTTTTGGTGGCTATGATGTATGCTGGAAGCTTCTCTTATGAGCACTTACTGCT 7297
Oy 781 GGATCAATAGAAGCTATTTGACCGATCCCTGCTTCTATATATCAGAGAAGATGCAAAAG 840
Db 7296 GGATCAATAGAAGCTATTTGACCGATCCCTGCTTCTATATATCAGAGAAGATGCAAAAG 7237
Oy 841 GGAAGAAAGCTCTAACTTAAATGATATGAAAGATAGCTTCTCTATCTTGTAGAT 900
Db 7236 GGAAGAAAGCTCTAACTTAAATGATATGAAAGATAGCTTCTCTATCTTGTAGAT 7177
Oy 901 CTCCTTATATCTTTTATTAATCTCTGTTGGTATGCGGATGATATGCTTATGCTTAA 960
Db 7176 CTCCTTATATCTTTTATTAATCTCTGTTGGTATGCGGATGATATGCTTATGCTTAA 7117
Oy 961 TCGAAGTACTTGAAGAAAGTCACTGAAAGTCAATATCCATATATATGATGACTATAGT 1020
Db 7116 TCGAAGTACTTGAAGAAAGTCACTGAAAGTCAATATCCATATATGATGACTATAGT 7057
Oy 1021 ACTTATGAGGAGCTTCTCTCTGAGCTGGCTGATGCTTATCTTATGCTATTTG 1080
Db 7056 ACTTATGAGGAGCTTCTCTCTGAGCTGGCTGATGCTTATCTTATGCTATTTG 6997
Oy 1081 TTGGTGTAGTCACTGATGCTTAAATTTGATGTTAACTGAGAGCCCTGACTCCGTCA 1140
Db 6996 TTGGTGTAGTCACTGATGCTTAAATTTGATGTTAACTGAGAGCCCTGACTCCGTCA 6937
Oy 1141 TGGTCTCTCTTACAGATATCGTTTCTGCTCTGTTATCTTATAGAAACCAAGCTTCTG 1200
Db 6936 TGGTCTCTCTTACAGATATCGTTTCTGCTCTGTTATCTTATAGAAACCAAGCTTCTG 6877
Oy 1201 GGGTGGTGGCTATGTTGGGTACAAGCTCTCTGATGCTGAGTGGTGTGCGAGCTATAC 1260
Db 6876 GGGTGGTGGCTATGTTGGGTACAAGCTCTCTGATGCTGAGTGGTGTGCGAGCTATAC 6817
Oy 1261 AGAATATCTCTTGCAGAAATCCACAAATACGCTCTCTTGGACTCAACTAAGAAATGGCT 1320
Db 6816 AGAATATCTCTTGCAGAAATCCACAAATACGCTCTCTTGGACTCAACTAAGAAATGGCT 6757
Oy 1321 ATATCCCTCTTGGACCAAGAGCAAAAGTCAAGGCTGCTATGATGATGCTGCG 1380
Db 6756 ATATCCCTCTTGGACCAAGAGCAAAAGTCAAGGCTGCTATGATGATGCTGCG 6697
Oy 1381 CCCGCTTGGGAAATCAGAGAGAGCTTATATCCACAAGGTTGCTGCTTATCTGTGAA 1440
Db 6696 CCCGCTTGGGAAATCAGAGAGAGCTTATATCCACAAGGTTGCTGCTTATCTGTGAA 6637
Oy 1441 GTATGGAGCTTGAAGCTTATCTTGGAGTATCTTCTTATCATATGCTATTTGGT 1500
Db 6636 GTATGGAGCTTGAAGCTTATCTTGGAGTATCTTCTTATCATATGCTATTTGGT 6577
Oy 1501 TGGTCTTGGCACTAAGTTAAACAACATATCTTGAAGCAAGTCTGCTTAAAGAACAAG 1560
Db 6576 TGGTCTTGGCACTAAGTTAAACAACATATCTTGAAGCAAGTCTGCTTAAAGAACAAG 6517
Oy 1561 AAGTGGCTCAGAGATTCAGTCTGCTGCTCTTCTATAGAGTGGCTTCTTACTTGT 1620
Db 6516 AAGTGGCTCAGAGATTCAGTCTGCTGCTCTTCTATAGAGTGGCTTCTTACTTGT 6457
Oy 1621 GATCCCTACCTGCTTTT 1637
Db 6456 GATCCCTACCTGCTTTT 6440

```

```

RESULT 2
AE001619 11421 bp DNA linear BCT 01-DEC-2000
LOCUS Chlamydia pneumoniae section 35 of 103 of the complete genome.
DEFINITION AE001619 AE001363
ACCESSION

```

```

VERSION AE001619.1 GI:4376620
KEYWORDS Chlamydia pneumoniae CWL029.
SOURCE Chlamydia pneumoniae CWL029.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 11421)
AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lamme, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W., and Stephens, R. S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 2 (bases 1 to 11421)
AUTHORS Kaiman, S., Mitchell, W., Marathe, R., Lamme, C., Fan, J., Olinger, L., Grimwood, J., Davis, R. W., and Stephens, R. S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
1. 11421
/organism="Chlamydia pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
complement(95..1069)
/gene="yfgD"
/note="Cpn0346"
complement(95..1069)
/gene="yfgD"
/codon_start=1
/transl_table=1
/product="Integral Membrane Protein"
/protein_id="AADI8490.1"
/db_xref="GI:4376621"
/transl_table="MALGSPDYGVSPFQSPVSPFSLFSGSLTSLYIDIOIIV
LAISGCAFGCTFLVRKMYANASHHTVLFGVCLTFTHOLTSLCTLLAAM
TAMTGLFLIRNTKRVSESSSTALVSLFSLVLVLPKNAITGELVGNAD
SLTKEDLPVATVILANAVITTFAPRSLVCSFDSVASSLGIPIRLVDYLIIFQUSA
SAYDLGSTGSISVFLTMVYIVKFIYFRGYSKNEKISEKSSQY"
complement(1069..2415)
/gene="yfgC"
/note="Cpn0347"
complement(1069..2415)
/gene="yfgC"
/codon_start=1
/transl_table=1
/product="Integral Membrane Protein"
/protein_id="AADI8491.1"
/db_xref="GI:4376622"
/transl_table="MLSCVPSDIFLSSFLAVTLICMTALMGTILLISKOPLESLS
SHASYPLVAGLMAQYVFSLOSIFWVILFGCAVSGYGIIVFLGKVKLHKDSAL
CFVYVFEAIGVILASVKSSEPLVIRINAYILGOAATGFEATLAALVFCSLPA
LMWYROIIVTTTPKDFAVMCGIKTVYKALSLIFISIVSGRSVCIYLSAMPFA
PSICARLSDRLSTILLISAFSGISALSYISVATCAITIGQAVPTLPGLY
VICGLIAGLCLFSPKSGWVIRPVRRKHSFSKDOEHLKVFMHISHNLENTSVD
FVCSKYQEVFGPKPFPFRWVOLIEMGWYVKEDDYRLTRKKGSEALRLVRHRLME
SYLVNSLDFSKESVHELAEIEHLTELDLTLEILNDCYDPRHOIIPNRKKEV"
complement(2418..3197)
/gene="yfgB.1"
/note="Cpn0348"
complement(2418..3197)
/gene="yfgB.1"
/codon_start=1
/transl_table=1
/product="rRNA methylase"
/protein_id="AADI8492.1"
/db_xref="GI:4376623"
/transl_table="MNVDKDEFMSVHNLVNYEHAVALYHISFLSGSLTALIGPNG
AGKSTLLKASLIGLTKPSSGIVYFENQFKKRVORIAVMPORASVDMFPPTVLDLIM
GCYSYKGMGRISDDRRAPHLIERVGLSVADROIGDSGGGOORAFILARALMOKA
DIVYIMDELFSALIMASFKTSYGVLOELRDGKTIYVYVHHLSHVRQLFDIVYLLNRL
ICCGPTDECLNGDTIIFQTYGCEIELELQTLKLSHGKQFGSC"
complement(3181..4164)

```

gene

```
gene
CDS
    /gene="ytga"
    /note="CPn0349"
    complement(3181..4164)
    /gene="ytga"
    /codon_start=1
    /transl_table=11
    /product="Solute protein Binding Family"
    /protein_id="AAD18493.1"
    /db_xref="GI:4376624"
    /translation="MDAKMGYIFKVMRIFCFVACGTCGCTNSGFQANSPCLLSM
    NRMTHDCVERVGNKRLATVLIKSLDHAHEWKGDKDTAGSAVIFCNLSGHEHTL
    SLRKLENNPNPSKGLERLARGAFPLEEDGICDPHTMDLSIKWEAVIETLEVLK
    KPEWSEAFKANSSELVCEMSILDSMAKOCSTIIPENLRYLVSGHNARSYTRRLAT
    PEVVASGARSRCISPEGLSPEADISVDIAVVDY INEHDSVYVPEPDTLNODALKK
    IVSSLKSHLVRLAQKRLYSDNDVNTSFTRKHNVCLTEELGVALECGR"
    4575..5081
    /gene="CPn0350"
    4575..5081
    /gene="CPn0350"
    /codon_start=1
    /transl_table=11
    /product="CT066 hypothetical protein"
    /protein_id="AAD18494.1"
    /db_xref="GI:4376625"
    /translation="MAVEOSHIEKEIEKLGKAIKRVGNKENDLCRYLPGSPGYMH
    HFTLKMSAPBOLLKMLKTFIESETPTRINPRAPRSGSKRDRIFNFKTDIER
    VLRLARQGDMDLARSFPRKPLSLKRELLRSTRNGIVSELMNAYEAKVASPN
    LEVYSPY"
    5282..6829
    /gene="adt_1"
    /note="CPn0351"
    5282..6829
    /gene="adt_1"
    /codon_start=1
    /transl_table=11
    /product="ADP/ATP translocase"
    /protein_id="AAD18495.1"
    /db_xref="GI:4376626"
    /translation="MTREKPKFGLRSFLMIPINTHELKRYLPFMEFCTTNTVL
    RDKDTLVGAPSGAEAIPIFKFLVVPCHITFMLYAKLSLISKQALVAVGTPF
    LIFLFPFVIVPLRDLHPTEFADRLQALLPGLGLVALIRNMTFAFVVLWLG
    SVMLSMGFANEITKTHEAKREFALGICANISLGSRAIYWASKRLARVSEGVG
    PWGISLRLLAMTIVSGVLNAYWIMKNVLTDPREFNPEMOKGKGAKPMKMD
    SFELARSPYILLALIVAYGICINLEVMWSOLKXYQNMNDYSEFMNFSPTG
    VSVYLMFVGVGNVIRKRGWLTGALVTPVWMLTGIVFEALVIRPNQSGIVAMGCTP
    PLMLAVYGAIGNILSKSTKALPDSITREKAYITPDQOKVGRADIVAPAREKSG
    GALIQGGLVIGSIGAMTPYLAVILFTIAIWLVSATKLNKLELAOSALEOVADE
    DSAPAS"
    6968..8227
    /gene="CPn0352"
    6968..8227
    /gene="CPn0352"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18496.1"
    /db_xref="GI:4376627"
    /translation="MKCSPLTVPHIIFLKNDCCHRSGLKIRTIARILISLVAYS
    ALSVFFLAARISVIGTLALAAIVILITTVVALAKSKVLPINELQKITINRYK
    EEFYFVKTHSLVNELEKIFINCMKSGDLPNLLKKAFAFGIDILKSIDLTFPEFEE
    ILONCPLYWLSHEIDKTESVAGEIGLKTQKVGLGLPLAFHNGYTTIFHSYRPL
    TLISEQKFLYSKASKNQMDSPVYKTCGEIPEKELPMNIFRKDOVQISQELFELS
    HGITWEOQMOLINPDNMKMLCOPDKAGKCEISMATPGCFINTEGMDPVSAYEPT
    VPMTKRELKYLLEKVESPMHPASALYOKICVMTTHONILKRMQVRNYSOWTSS
    LPOYAFHQIYKLEKLISSLPTRSSL"
    8287..8532
    /gene="CPn0353"
    8287..8532
    /gene="CPn0353"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18497.1"
```

```

    /db_xref="GI:4376628"
    /translation="MRPNIKSLIPIKIRYSSQKPOKARPEFKKSINSLVILL
    LEAFENPSSIIQNFKNKRNKNINISIRLIVVFTT"
    8561..9904
    /gene="CPn0354"
    8561..9904
    /gene="CPn0354"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18498.1"
    /db_xref="GI:4376629"
    /translation="MSNTSPVIONRSCNYPELKNSTTHIYISATILGALIAFL
    CVAPVSYIGALLGGLITALLIGVILGKIKITPMISKREVPQDELVNRIRAHYK
    FVSDFVSARKNLNDLISFTDLLQNLSEVSSSTRNYSEELQKITFDEBIARUKNE
    VRTASLRLESASRRLEPSLPKTILOKVPFFPMGEIISAGSVVELIHRVKITIGSL
    EEDLSDYIKRPMELPYWLIPLDFRPTNSIINLTPLVRLATVRLDPOHLKALNGE
    WNLNHSDLNTPKQOLFARKYAAQYOSKHLSPQSDQEDFVNLICIFKHRYSMKQMSL
    IKTVPADLMENLCITLDPTRQPMDEPASFILGTYTQGLIHKSEEARLISLTLISD
    QFKTRRQSTNIAMFLEMLATHNSFRSLRPIITYHPLKRSVFSQPEDESSLIG"
    complement(9982..11283)
    /gene="CPn0355"
    complement(9982..11283)
    /gene="CPn0355"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAD18499.1"
    /db_xref="GI:4376630"
    /translation="MSMTIVPHALFKNHCESHSTFPLSSRTIVRLIASLFCIGALAA
    LGCLAPVSYIVSYLAFVILISVILALIFGKKKLPPTRIIPDRFTVIDEAYG
    LSISAFVREDOVTLAEFRPSTALLCNISPEEIKQOLPSELRSKVESGISTRAGDLE
    KNMRIFEDLSQCPILYWLQKFTSAGDPQVCRDGVPRCEGYIWLGLPYSTAKAT
    IFCKETHHILQOLIKREVDLLKNALEKWPDEKALVERIYTYTYARGLKTEAGG
    LTKETISKEILLSLHGSFQDQILQILPDPADMDICFVNSYAVNIQCALVGLALS
    SONLDESSIDFVNLGLYVIODLKEAVQASADEPKELGKFLRLHSVSKRLES
    VLROGLIHAIHEHNAARVYDVNPTVGARLHHRRTSIFFXD"
BASE COUNT      3791 a      2376 c      1936 g      3318 t
ORIGIN
Query Match      90.7%; Score 1484; DB 1; Length 11421;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      1  GAAATAAAAACTATCAGATATGAAATATAAAGTATTTTCAGAGGTAATATGACAAAA 60
      |||||||
Db      5232  GAAATAAAAAACTATCAGATATGAAATATAAAGTATTTTCAGAGGTAATATGACAAAA 5291
QY      61  CCGAAGAAAAACCTTTGGAAATATGCGCTCTTCTGTGGCCGATPACTACTCAGAGC 120
      |||||||
Db      5292  CCGAAGAAAAACCTTTGGAAATATGCGCTCTTCTGTGGCCGATPACTACTCAGAGC 5351
QY      121  TAAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGATTAACATTTAATAGGTTG 180
      |||||||
Db      5352  TAAAGAAAGTTCTGCCAATGTTCTTAATGTTCTTCTGATTAACATTTAATAGGTTG 5411
QY      181  TACGGGATACAAAGACATCTTATTGGGAGGCTCGGTCTGGTGCAGAGGCAATAC 240
      |||||||
Db      5412  TACGGGATACAAAGACATCTTATTGGGAGGCTCGGTCTGGTGCAGAGGCAATAC 5471
QY      241  CTTTCATCAAGTTTGGCTGTGTGCCCTGTGCGCATTAATCTTATAGCTATTATGCAA 300
      |||||||
Db      5472  CTTTCATCAAGTTTGGCTGTGTGCCCTGTGCGCATTAATCTTATAGCTATTATGCAA 5531
QY      301  AGCTAAGTAATATTTTAACTAAGACAGGCCCTATTTTATGACAGTGGAAAGCCCTTTTAA 360
      |||||||
Db      5533  ACCTAAGTAATATTTTAACTAAGACAGGCCCTATTTTATGACAGTGGAAAGCCCTTTTAA 5591
QY      361  TTTTCTTTGCCCTGTTCAGCATTAATTTATCCGCTACAGCATTTTATCACTCTACAG 420
      |||||||
Db      5592  TTTTCTTTGCCCTGTTCAGCATTAATTTATCCGCTACAGCATTTTATCACTCTACAG 5651
QY      421  AATTGTGTCAGCGTTTACAGGCATCTACCTCCAGGATTTCTTAGACATCGTTGCCATCT 480
```

```

|||||
Db 5652 AATTGCTGACCGTTTACAGGCGCTACCTCCAGAGATTGCTAGACCTCGTTGCCACT 5711
Oy 481 TAAGAACTGAGCATTTCTGATCTTTTATGTAAGTGTGAGAACTATGAGGAAAGCCGATGC 540
Db 5712 TAAGAACTGAGCATTTCTGATCTTTTATGTAAGTGTGAGAACTATGAGGAAAGCCGATGC 5771
Oy 541 TATCTATATGTTCTGGGATTTGCTAATGAATTAACAAAATCCAGCAAGCAAGCGCT 600
Db 5772 TATCTATATGTTCTGGGATTTGCTAATGAATTAACAAAATCCAGCAAGCAAGCGCT 5831
Oy 601 TCTACGCTCTTTTGGGATTCGAGCTAATATTTCTTTACTAGCTCTGCTGCTGCAATTG 660
Db 5832 TCTACGCTCTTTTGGGATTCGAGCTAATATTTCTTTACTAGCTCTGCTGCTGCAATTG 5891
Oy 661 TTTGGGCTTCAAGTTGAGAGCTTCGCTTCTGAGAGGTGATGATCCTTGGGAAATTTCT 720
Db 5892 TTTGGGCTTCAAGTTGAGAGCTTCGCTTCTGAGAGGTGATGATCCTTGGGAAATTTCT 5951
Oy 721 TACGCTTTTGTATGCTATGATATGATGAGACTTGTCTTATGAGCCAGTTACTGCT 780
Db 5952 TACGCTTTTGTATGCTATGATATGATGAGACTTGTCTTATGAGCCAGTTACTGCT 6011
Oy 781 GGATCAATTAAGACGATTTGACCGATCCTCGCTTCTAATATCCAGAGAATGCAAAAG 840
Db 6012 GGATCAATTAAGACGATTTGACCGATCCTCGCTTCTAATATCCAGAGAATGCAAAAG 6071
Oy 841 GGAAGAAAGGCTTAACCTAAATGATGAAAGATGAGCTTCTCTATCTGTTAGAT 900
Db 6072 GGAAGAAAGGCTTAACCTAAATGATGAAAGATGAGCTTCTCTATCTGTTAGAT 6131
Oy 901 CTCCTTATATCTTTTATTAACCTCTCTGTTATGCTTATGCTATGTTGCAATTAACCTTAA 960
Db 6132 CTCCTTATATCTTTTATTAACCTCTCTGTTATGCTTATGCTATGTTGCAATTAACCTTAA 6191
Oy 961 TCGAAGTACTTTGCAAAAGTCAAGCTGAAAGCTGCAATATCTTATATGATGACTATAGTG 1020
Db 6192 TCGAAGTACTTTGCAAAAGTCAAGCTGAAAGCTGCAATATCTTATATGATGACTATAGTG 6251
Oy 1021 AGTTGATGGGAACTTCCTCTGAGTGGGATGATGCTGATTCATATGCAATTTTG 1080
Db 6252 AGTTGATGGGAACTTCCTCTGAGTGGGATGATGCTGATTCATATGCAATTTTG 6311
Oy 1081 TTGCTGATTAAGCTATTCGTAATTTGATGATGATGAGCCCTAGTCACTCTGCA 1140
Db 6312 TTGCTGATTAAGCTATTCGTAATTTGATGATGATGAGCCCTAGTCACTCTGCA 6371
Oy 1141 TGGTTCTCTTAACAGGATTCGTTTCTGCTTCTGTTATCTTTAGAAACCAAGCTTCTG 1200
Db 6372 TGGTTCTCTTAACAGGATTCGTTTCTGCTTCTGTTATCTTTAGAAACCAAGCTTCTG 6431
Oy 1201 GAGTGTGCTATGTTGCTGCAACTCTCTCATGCTAGCTGTGTTGCGAGCTATAG 1260
Db 6432 GAGTGTGCTATGTTGCTGCAACTCTCTCATGCTAGCTGTGTTGCGAGCTATAG 6491
Oy 1261 AGAATATTTCTTTCGAAATCCACAAAATACGCTCTCTTGTAGCTCAACTAAGAATAGGCT 1320
Db 6492 AGAATATTTCTTTCGAAATCCACAAAATACGCTCTCTTGTAGCTCAACTAAGAATAGGCT 6551
Oy 1321 ATATCCCTCTTGACCAAGCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 1380
Db 6552 ATATCCCTCTTGACCAAGCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCA 6611
Oy 1381 CCGGCTTGGGAAATCAGAGAGGCTTTAATCCAAAGGTTGCTGTTATCTGTGAA 1440
Db 6612 CCGGCTTGGGAAATCAGAGAGGCTTTAATCCAAAGGTTGCTGTTATCTGTGAA 6671
Oy 1441 GTATTGAGCTATGACCCCTTATCTTTCGATGATTTCTTTTCAATTCATTTGCTATTTGCT 1500
Db 6672 GTATTGAGCTATGACCCCTTATCTTTCGATGATTTCTTTTCAATTCATTTGCTATTTGCT 6731
Oy 1501 TGGTTTCTGCAACTAAGTTAAACAACATATTTCTTATGCGCAGCTCTGCTTTAAGAACAG 1560
|||||

```

```

Db 6732 TGGTTTCTGCAACTAAGTTAAACAACAAATATCTTACGCCACTGCTGCTTTAAGAACAAG 6791
Oy 1561 AAGTGCTCAAGAGATTCACGCTCCTCTCTCTATAGTATGCTTCTTACTCTTGT 1620
Db 6792 AAGTGCTCAAGAGATTCACGCTCCTCTCTCTATAGTATGCTTCTTACTCTTGT 6851
Oy 1621 GATCCCTACCTGCTTTT 1637
Db 6852 GATCCCTACCTGCTTTT 6868

RESULT 3
AP002546
LOCUS
DEFINITION
Chlamydia pneumoniae J138 genomic DNA, complete sequence,
section 2/4
ACCESSION
AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801
AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808
AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815
AB036079 AB036080 AB036081 AB036082 AB038348 AB038349 BA000008
AP002546.2 GI:10176692
VERSION
KEYWORDS
SOURCE
ORGANISM
Chlamydia pneumoniae J138 (strain:J138) DNA.
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
1
Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,
Takeshima,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,
Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,T.,
Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T.
Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and
the United States
J. Infect. Dis. 181 suppl 3, S524-S527 (2000)
20298986
JOURNAL
MEDLINE
REFERENCE
2
Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CML029 from USA
Nucleic Acids Res. 28 (12), 2311-2314 (2000)
20330349
JOURNAL
MEDLINE
REFERENCE
3
(bases 1 to 300650)
Shirai,M.
Direct Submission
Submitted (04-JUL-2000) Mitsunori Shirai, Yamaguchi University
School of Medicine, Department of Microbiology, 1-1-1
Minamikogushi, Ube, Yamaguchi 755-8505, Japan
(E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,
Fax:81-836-22-2415)
On or before Sep 15, 2000 this sequence version replaced
gi:6172290, gi:6172292, gi:6172294, gi:6172296, gi:6172326,
gi:6172328, gi:6172330, gi:6172332, gi:6172334, gi:6172336,
gi:6172338, gi:6172382, gi:6172384, gi:6172386, gi:6172388,
gi:6172390, gi:6172392, gi:6172394, gi:6174666, gi:6174668,
gi:6635174, gi:6635176, gi:6635178, gi:6635180, gi:8547433,
gi:8547438, gi:8978640.
Location/Qualifiers
1..300650
/organism="Chlamydia pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
144..935
/gene="CPj0267"
144..935
/gene="CPj0267"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BA08477.1"
/db_xref="GI:8978641"
/translation="MSIMSLNKTNALNLPPEAVCLNAMPKXYINDRKTFACTVTL

```



```

gene
957..1343
/gene="CpJ0268"
CDS
957..1343
/gene="CpJ0268"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA098478.1"
/db_xref="GI:8978642"
/translation="MSOCSSSTSWEMKSEFVPMKNKPTPLSPIPSEDEFILAYEP
FVLKTPDENANANPGSTSTPNVENGIDDLNPLGGPNEONNANNPGTSSGNPSLIPA
PERLPETENSOEEEGSQNNEDLG"
complement(1495..2472)
/gene="CpJ0269"
CDS
complement(1495..2472)
/gene="CpJ0269"
/codon_start=1
/transl_table=11
/product="olipeptidase"
/protein_id="BA098479.1"
/db_xref="GI:8978643"
/translation="MTIDHCDLLSHPHCRKDPVAVRCSPEOLLSGGROOVCAIFV
HSKSEPNCDKONSLEFSLPNQYPIDGLISYEERENGSSQKSLSLNSTENASALGP
DTAPLGLTLAKLILHTKQGPLAYLGIVKNGNRGQTEAPKRLSNDKVLDDIMYEL
GVPIDLSHCOSKLAIEDIIDYADKLPNLAIASHNSFVSVDHNRNLDAHAKEIVKR
KGVLGNLVRSYVGDSDLDLEKHVILHNLGLISIVSGDFEYANEDENEFNECCSS
AAHPVLNOLIHRIFFSKAESILSRAEKLKQIVQVNPVKITDVKL"
complement(2487..3347)
/gene="ywc"
CDS
complement(2487..3347)
/gene="ywc"
/codon_start=1
/transl_table=11
/product="Suas superfamily-related protein"
/protein_id="BA098480.1"
/db_xref="GI:8978644"
/translation="MPDKKAQITFSLPVMSAIIHQKIVALPTDVTYGVFLVYASEA
BERLYALKDRPSKAFALYVNSIEDIENISGYPLPTKKLAQLFPGAITLVVHRNP
RPFETLAFRIIVDSVRAEIVDHCGTILIGTSANLSEPSALTAQIEFADPDHDKICF
DPCSHGLESTIVASDPLIYIREGLISRSVLEINAGTEAKIFHRTSHAFSKHIKYIV
KNOBOLVFLSGSLDFKGVCEHPRKPFYTRLRKALKKRTPSYIFTIDINTSDPEL
PFLSPYITP"
3656..4387
/gene="CpJ0271"
CDS
3656..4387
/gene="CpJ0271"
/codon_start=1
/transl_table=11
/product="lysophospholipase esterase"
/protein_id="BA098481.1"
/db_xref="GI:8978645"
/translation="WTOXSFFRRKIGNTEATECRNPDDPIITIIICHGCSLADNTEP
PSICSFSLRPWTWIFPNKILPLENDPFRGSRCAFLNVLLOELSLYANGNIOEKY
DELFDVDLETEREALELILNLRPNYNETIIGGFSQGAIIATHVLVTSQNPYAGLIF
AGARLFGNGWEGEELKQCAQVPELOSHEDEILPVLHGAIHLDLLTLKNGOFVSFHG
GHEIPSYVFOKQVTVPMNIDPARC"
complement(4359..5255)
/gene="dnaX_2"
CDS
complement(4359..5255)
/gene="dnaX_2"
/codon_start=1
/transl_table=11
/product="DNA polymerase III gamma and tau"
/protein_id="BA098482.1"
/db_xref="GI:8978646"
/translation="WHEENQGWELARKVYHOEVPPAILLHGCTPLIDOKRAEQLA
SEILSSPSGEHVKVOKIHPIYOFPPEGGRSLSIDLPKIKQIYISFEAVYKI
YIIIEADRMTLAISAFLKVEEPKHAVALIITLTAQVRLPKTIISRLSIFIEGKEK
ILCKETFSYLFRAQCEIPLVTEVSQIIKESSETDKQVLRDKVQRFEMVLLLELYDRY

```

```

gene
complement(5246..5866)
/gene="tdk"
CDS
complement(5246..5866)
/gene="tdk"
/codon_start=1
/transl_table=11
/product="thymidylate kinase"
/protein_id="BA098483.1"
/db_xref="GI:8978647"
/translation="MFIVIEGEGSGKSLAKALGDQVLAODRKVLLTRREGCLIGE
RLRDLILEPHELSRCELEFLFGSRAQHOEVIIPALRGQYIVICERFDSTIVQ
GIAREGADFADICSKVSGTPEPLPNVILLDDIPADIGLORKHROKVEDEKPKLS
VHNRIREFLSLASDPSRYIVLARESLASLIDKVMILHTOLGCT"
complement(5871..8375)
/gene="gyrA_1"
CDS
complement(5871..8375)
/gene="gyrA_1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit A"
/protein_id="BA098484.1"
/db_xref="GI:8978648"
/translation="MPNKDEIIVRKNLEEMKESYLARSMSVITSRALPIDRIGKPS
ORVLYAMKOLISLSPGAKHRCACIKGTSDDYRPHGSESVIYPLVRAQWAMARYPL
VDGQNGFESIDGDEPPAARKYIEARLTHSAMLMEDLDQVDVDPNDETIEHPVFP
SKFPLNCGSSGGLAVGMAVNIIPHNILGELIATLLANQASVDELQWHPEDF
TGCIGISGEGIRSTYTGTRGKIKVRAARLHVENEDKHRESIITTEMVYNKSRILEQ
IANVNEKTLISDVREDSQGRIVALEIKGSSSTIINRLYKFPDVOVTFGANG
LADKNPRTMSHMTSAMI RHRREVTRRTREYELNKAETRAVLEGYLAKALCLDA
LVKTIRESKNEHAKERTIESFGFTPEPALAILERILYQLTGLEAEKIQOKYBELINK
IAYTKQVLSDEGLVKDITRNELDLKLHKVARRTTEFDADRIDIEDIITNESVIT
TISGDVYKRMVAVVFEKQRRGGHGVGFDKKAGFLKAVYSATKQYLLIFNPGQ
CYWLKVMQLPEGERAKGKPIINFLGEGIRPEELAAIINIKNPNAGFLPLATRGVY
KKSILDAFNSPRKKGIRALEIDEGBELAAHIVSDEEKVMLFTHLGMAVREPHKVR
PMGRARGVRCVSLKNEEDKYVSCOIYENOSVLIYCDQGGKRSIVDERETRGVY
GVRSLIERNGNVIGATIPVTDHOSIILMSSQGAIRINMDVYVMKSGTQVRLVHL
KEGDALVMEKLSNENDEVLSGSEECSDTVSLR"
complement(8391..10808)
/gene="gyrB_1"
CDS
complement(8391..10808)
/gene="gyrB_1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BA098485.1"
/db_xref="GI:8978649"
/translation="MDPEKKNYDASATIVLGLQAVRPGMYIGDNGITGLHLVVE
VVDNSIDPAMAGYCSRIDVRILEDQGIYIVNGRGIPREVERESANQGRVSALEY
LTVLHAGKEDKDSIKVSGGLHGVGCVNALSEKLVATVFKDKKCIOMERSKGIPTV
PLQVYVSDRQGTETVEYDPDKIFSTCFEDSILMKRLRELAFINRGITVFEEDRDV
SFDKVTFPEBGIOISFVSYLNQNKSEISFPIYICGTVGDDGEIEFEPAALQMSGYS
ELVYSYANNIPTRQGGTHLGEFSTALPTVIMYIKRANHLAKNNKLLALGEGIRGLTA
VISVVPNPQFQGTOKKLGNSDVSAVOQVGAETITFEFENQIAMIYDKVAVAA
QAREAKARBELTLKRSKLDSDARLPKGIIDLENDPECEKNIYBGDSAGSAAQGRD
RRFOALITPRKTLINVEKARLQKIFQNEICGTITIAACGCGADPNKSKRYRRIIT
MTDADVDSSHRTLLTFYFHYMHALINENCVYIAQPLVYVSKKERYLLISEKEMD
SYLLMLGTNESIILFKSTERLRCGALSEFINVILDVSEFINTELEKKAIPSEFLEWY
KEGIGYPLVYLAPTQMGQGRIVYISDEKEEALAOEHTKRIITELXYVVAEVIQNG
LKEGYGLDISYVILQKNEIYVIGNEPSCNYSCTLEVINYKINLGRKGIETQRYGQ
LGEHMADQLMPTTANPQORTLIHVSLSKDAVEADHITFLMKGEVPPREFTESHALSI
RINNLDT"
complement(10812..11159)
/gene="CpJ0276"
CDS
complement(10812..11159)
/gene="CpJ0276"

```

```

Query Match 90.7%; Score 1484; DB 1; Length 300650;
Best local Similarity 99.8%; Pred. No. 0;
Matches 1634; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
0Y 1 GAAATTAAGAAACGATTCAGAAATGAAGTATTTTCAGAGCGTAAATATGACAAAAA 60

```

```

|||||
Db 93946 GAAATTAATAAATATCAGATGAGAAATAAAGTATTTAGAGGGTAAATATGACAAAA 94005
Oy 61 CCGAAGAAAAACCTTTTGGAAAAATTTGCGCTCTTTCTTTGGCGGATACATACACAGAC 120
Db 94006 CCGAAGAAAAACCTTTTGGAAAAATTTGCGCTCTTTCTTTGGCGGATACATACACAGAC 94065
Oy 121 TAAAGAAATCTTCCCAATGTTCTTAATGTTCTTCTATTAACATTTAACTATACGGTGT 180
Db 94066 TAAAGAAATCTTCCCAATGTTCTTAATGTTCTTCTATTAACATTTAACTATACGGTGT 94125
Oy 181 TACGGGATACAAAAGACACTCTTATTTGTGGAGCTCCTGGTCTGGTGCAGAGCAATAC 240
Db 94126 TACGGGATACAAAAGACACTCTTATTTGTGGAGCTCCTGGTCTGGTGCAGAGCAATAC 94185
Oy 241 CTCTTATCAAGTTTGGCTGTGTTGCCCTGAGCTATTAATCTTTATAGCTATTTATGCAA 300
Db 94186 CTCTTATCAAGTTTGGCTGTGTTGCCCTGAGCTATTAATCTTTATAGCTATTTATGCAA 94245
Oy 301 ACCTAAGTAAATATTTTAACTAAGACAGGCTTATTTTATGACAGTGGAGACGCCCTTTTAA 360
Db 94246 ACCTAAGTAAATATTTTAACTAAGACAGGCTTATTTTATGACAGTGGAGACGCCCTTTTAA 94305
Oy 361 TTTTCTTGGCCCTGTTCCCGACTGAATTTATCCGCTAGCGCATGTTTACATCCGACAG 420
Db 94306 TTTTCTTGGCCCTGTTCCCGACTGAATTTATCCGCTAGCGCATGTTTACATCCGACAG 94365
Oy 421 AATTTGCTGACCGTTTACAGGCCATTCCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 480
Db 94366 AATTTGCTGACCGTTTACAGGCCATTCCTACCTCAGAGATTGCTAGAGCTGTTGCCATCT 94425
Oy 481 TTAAGAACTGGACATTTTGTCTGATTTTATGTACTTGTCTAAGTATGAGGAGCGTCAATGC 540
Db 94426 TTAAGAACTGGACATTTGCTGATTTTATGTACTTGTCTAAGTATGAGGAGCGTCAATGC 94485
Oy 541 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAATTCACAGAGAAAGCGTGT 600
Db 94486 TATCTCTAATGTTCTGGGATTTGCTAATGAATTAACAAAATTCACAGAGAAAGCGTGT 94545
Oy 601 TCTAGCGCTTTTTCGATTCGAGGCTAATATTTCTTTACTAGCTCTGTGCTGCAATTG 660
Db 94546 TCTAGCGCTTTTTCGATTCGAGGCTAATATTTCTTTACTAGCTCTGTGCTGCAATTG 94605
Oy 661 TTTGGGCTTCAAGTTGAGAGCTTCCGTTTCTGAAGGTGTAGACTCTGGGGAATTTCTT 720
Db 94606 TTTGGGCTTCAAGTTGAGAGCTTCCGTTTCTGAAGGTGTAGACTCTGGGGAATTTCTT 94665
Oy 721 TACGCTTTTGTAGTGGCTATGACTATTTGATCTGAGACTGTCTTATGCGCAATTTCTGCT 780
Db 94666 TACGCTTTTGTAGTGGCTATGACTATTTGATCTGAGACTGTCTTATGCGCAATTTCTGCT 94725
Oy 781 GGATCAATAGAACTATTTGACCGATTCCTCGCTTCTATTAATCCAGAGAAATGCAAAAGG 840
Db 94726 GGATCAATAGAACTATTTGACCGATTCCTCGCTTCTATTAATCCAGAGAAATGCAAAAGG 94785
Oy 841 GGAAGAAAGGTGCTTAACCTTAATGAATGATGAAGATAGCTTCTCTGATCTTGTAGAT 900
Db 94786 GGAAGAAAGGTGCTTAACCTTAATGAATGATGAAGATAGCTTCTCTGATCTTGTAGAT 94845
Oy 901 CTCCTTATATCTTTTATTAATCTCTCTGATGATGCGATGCTATTTGCAATTAACCTAA 960
Db 94846 CTCCTTATATCTTTTATTAATCTCTCTGATGATGCGATGCTATTTGCAATTAACCTAA 94905
Oy 961 TCGAGTACTTGGAAAGTACAGTGAACATGCAATATCTCAATATGATGATGACTATAGTG 1020
Db 94906 TCGAGTACTTGGAAAGTACAGTGAACATGCAATATCTCAATATGATGATGACTATAGTG 94965
Oy 1021 AGTTGATGGGGAACCTTCTCTCTGAGCTGGCGTATGATCCGTAATATCATGCTAATTTG 1080
Db 94966 AGTTGATGGGGAACCTTCTCTCTGAGCTGGCGTATGATCCGTAATATCATGCTAATTTG 95025
Oy 1081 TTGGGTGAAGCTATTTGCTAAATTTGATGATGATTAACGTGAGACCTATCACTCTGTGCA 1140
|||||

```

```

Db 95026 TTGGGTGAAGCTATTCGTAAATTTGGATGCTTAACGTGAGACCTAGTCACTCTGTCA 95085
Oy 1141 TGGTCTCTCTAACAGATATCGTTTCTTCCGCTCTGTTATCTTTTGAACCAAGCTTCTG 1200
Db 95086 TGGTCTCTCTAACAGATATCGTTTCTTCCGCTCTGTTATCTTTTGAACCAAGCTTCTG 95145
Oy 1201 GCGTGGTGGCTATGTTGGTGTACACCTCTCTCATGCTAGCTGGTGTGTGGAGCTATAG 1260
Db 95146 GCGTGGTGGCTATGTTGGTGTACACCTCTCTCATGCTAGCTGGTGTGTGGAGCTATAG 95205
Oy 1261 AGAATATCTTCTGCAAAATCCACAAAATAGCTCTCTTGACATCACTAAAGAAATGGCT 1320
Db 95206 AGAATATCTTCTGCAAAATCCACAAAATAGCTCTCTTGACATCACTAAAGAAATGGCT 95265
Oy 1321 ATATCCCTCTTACCAAGAGCAAAAAGTCAAAAGTAAAGCTCTATTAAGTATGTCGG 1380
Db 95266 ATATCCCTCTTACCAAGAGCAAAAAGTCAAAAGTAAAGCTCTATTAAGTATGTCGG 95325
Oy 1381 CCCGCTTGGAAAATCAGAGAGAGCTTTAATCCACAAGATTTGCTCTATCTGTGAA 1440
Db 95326 CCCGCTTGGAAAATCAGAGAGAGCTTTAATCCACAAGATTTGCTCTATCTGTGAA 95385
Oy 1441 GATGAGAGCTATGACCCCTTATCTTGACAGTATCTTCTTCAATGCTATTTGGT 1500
Db 95386 GATGAGAGCTATGACCCCTTATCTTGACAGTATCTTCTTCAATGCTATTTGGT 95445
Oy 1501 TGGTCTTGCACACTAAGTTAAACAACACTATTTCTTACGACAGTCTCTCTTAAAGAACAAG 1560
Db 95446 TGGTCTTGCACACTAAGTTAAACAACACTATTTCTTACGACAGTCTCTCTTAAAGAACAAG 95505
Oy 1561 AAGTGGCTCAGAGATTTACGCTCTGCTCTTCTCAATAGTTGCTTCTTACTCTTGT 1620
Db 95506 AAGTGGCTCAGAGATTTACGCTCTGCTCTTCTCAATAGTTGCTTCTTACTCTTGT 95565
Oy 1621 GATCCCTACCTGCTTGT 1637
Db 95566 GATCCCTACCTGCTTGT 95582
|||||

RESULT 4
AE002301/c 23773 bp DNA linear BCT 26-MAY-2000
LOCUS Chlamydia muridarum, section 33 of 85 of the complete genome.
DEFINITION AE002301 AE002302 AE002160
ACCESSION AE002301.2 GI:8163203
VERSION
KEYWORDS
SOURCE
ORGANISM Chlamydia muridarum.
Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
REFERENCE
1 (bases 1 to 23773)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 20150255
PUBMED 10684935
REFERENCE
2 (bases 1 to 23773)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On or before Jun 1, 2000 this sequence version replaced gi:7190382,
gi:7190372.
COMMENT
FEATURES
Location/Qualifiers
Source 1..23773

```

gene  
CDS /organism="Chlamydia muridarum"  
/db\_xref="taxon:83560"  
/note="synonym: Chlamydia trachomatis MoPn"  
106..1842  
/gene="TC0330"  
106..1842  
/gene="TC0330"  
/note="protein export protein, FHIPEP family; identified by match to Pfam protein family HMM PF00771"  
/codon\_start=1  
/transl\_table=11  
/product="protein export protein, FHIPEP family"  
/protein\_id="AAF73547.1"  
/db\_xref="GI:8163204"  
/translation="MNHIRGSSMKIVTPICLITLVIPLRWIDPGVCANLCSHS  
ILFWVESRSSASARIFFPSLLIYLICLRILGLINASTFMTILASGWASPLIFALSFSLS  
GISPAVLVLCILPRINFILYTKGAERAEVRAFRSILESLDPKMSLDIAISAGRIS  
SRASRKNSLEESDYEFAMEGVFRRFKGDAIKMSVLLGVNLSIALFLGNHVGYNIYM  
LTVIDADAVSQPALITSCSAATLIAAKGEESLAOHVEYEEOGRQSFYIALIIQG  
MAFIIGARKPALIGFSVILLGLGYSPTYSEESLDFEKKERIELLPKNKENPVSLYRAA  
RHIOIUEJGFVEGESVVRVYNVSPRLVFSGEISLTELCSAMLKAMKRLPERTGG  
EFFRIALIEFPFOEGGSIEETIPIKISENSIVFLRALVKEOSLIHLPKITLAIIDL  
FGSQAKSSELVEVSRYNLVGQIGLSIMNRKDYLEVTIVBSLVQFPRDQAKEYFDNL  
NEKVSVQKDLREBGENFAIVTGSETRKLRTIVDPPEPLLVIHAHSELPMEPTT  
LNGAVSDVELLS"  
1950..2711  
/gene="TC0331"  
1950..2711  
/gene="TC0331"  
/note="similar to GB:X68709 PID:47122; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="RNA polymerase sigma factor, sigma-70 family"  
/protein\_id="AAF39194.1"  
/db\_xref="GI:7190373"  
/translation="MKTODLADTWOLYWSTKEIHHRDFLLIYAULPYKANVAHRLASGM  
PSHAVIEDLVASGVGLVRAVEREDPEKSRSFEYALFIIKAITIDLRQDWPRS  
YQVARRLADAMDALROTGLKEPTDGDCLEYINISOELSHMFSSRALVLSINDPS  
CODEGEGALTEERIADERAENGVDYIRKKAISTITEALLSLDEKERQWVALYYDDL  
VKEIGKTLGVSSESRSVSIHKALKILRGITLSLL"  
2869..4107  
/gene="TC0332"  
2869..4107  
/gene="TC0332"  
/note="similar to SP:P00952; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="tyrosyl-tRNA synthetase"  
/protein\_id="AAF39195.1"  
/db\_xref="GI:7190374"  
/translation="MQQLIDSLQKRGIIDNSSAGLESILAPVASAYLGCFPPAPSLHTIG  
FMHIGCEIFRLISAVGITPIALVGTATGMIGDSPGKSYERSLLDOEVLNDSKIEVAL  
ANYEDITIVNNADMIGSLMWDELRIDIGKYFRGSMIAKDVQARYSEEGISLYTES  
SYLLIOSDFVAILFKHNHGVNLOCGSDQMWTGICDIYIRKGIGQAGFGTYLPIDS  
SKCKIAGKESGVTVMLDEPLTSPYELFQFTLRSOEPKIAIKRMITLLDDPEVALDKR  
LENDQOAQRVYAAEVYKDVGAECILAOALTTSFPANKNKITBELDALIVOSGCG  
INVAADVIYGRMLDVVQDLGFCSSKGEARRLIDQRLGVNQEPILIDESVLDGTYLC  
PDRIYLLISQGKKKKRQVIDLN"  
4117..5556  
/gene="TC0333"  
4117..5556  
/gene="TC0333"  
/note="similar to SP:P00349 PID:1193; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="6-phosphogluconate dehydrogenase,  
decarboxylating"  
/protein\_id="AAF39196.1"  
/db\_xref="GI:7190375"  
/translation="MAPADIIGLIGLAVMGKNLVLNMIDHGFAVSVNRSRDEKTEEFLK

[illegible]

```
gene
CDS
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAF39200.1"
    /db_xref="GI:7190379"
    /translation="MQTNLENNLKHFIFCCPNQKISKMRLLRKKQKPL"
    /gene="TC0338"
    /note="TC0338"
    /note="similar to PID:790546 PID:1777933 GB:AE000520; identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="ABC transporter, periplasmic substrate-binding protein, putative"
    /protein_id="AAF39201.1"
    /db_xref="GI:7190380"
    /translation="MEFLHVRKRYKHVIGLLICAGCFYISSSGSRGKKSIDERTHITL
    SMNRITVCVRITGDRKRVLYLDGSDTPHSYEMVKDEDEMAISQILFCGLGLEH
    SASLRKHLGNSKVIDLGARLIDKNCVLESDPEPDPIHWDMGWISYKEMASVL
    VQOIPQVAEEFQKNAEQILSEMEDLRMAVSLAIPKRNRYLYTGHNAFSYFTRRYL
    SSDEERESGNMKLRCSPEGLSPEAQISRIDMRVEXICANDVGVFLIEDTLNDAL

Query Match
Best Local Similarity 100.0%; Pred. No. 3.5e-13; Length 23773;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 GAGCTAAAGAAGTCTCGCAATGTTCTTAATGTTCTCTGATT 161
DB 9140 GAGCTAAAGAAGTCTCGCAATGTTCTTAATGTTCTCTGATT 9096

RESULT 5
TAJ10586 1587 bp DNA linear BCT 23-MAR-1999
LOCUS Chlamydia trachomatis npt1 gene.
ACCESSION AJ010586
VERSION AJ010586.1 GI:4538859
KEYWORDS npt1 gene; nucleoside triphosphate transport protein 1.
SOURCE Chlamydia trachomatis.
ORGANISM Chlamydia trachomatis.
REFERENCE 1 (bases 1 to 1587)
AUTHORS Tjaden,J., Winkler,H.H., Schwoppe,C., Van Der Laan,M., Mohlmann,T.
and Neuhaus,H.E.
Two nucleotide transport proteins in Chlamydia trachomatis, one for
energy
Two nucleoside triphosphate uptake and the other for transport of
energy
JOURNAL J. Bacteriol. 181 (4), 1196-1202 (1999)
MEDLINE 99138740
PUBMED 9973346
TITLE 2 (bases 1 to 1587)
AUTHORS Neuhaus,E.
JOURNAL Direct Submission
TITLE Submitted (25-AUG-1998) Neuhaus E., University of Osnabrueck,
Biology / Plant Physiology, Barbarastr.11, D-49069 Osnabrueck, FRG
FEATURES
source
    /organism="Chlamydia trachomatis"
    /db_xref="taxon:813"
    /gene="npt1"
    /gene="npt1"
    /gene="npt1"
    /function="ATP/ADP transport protein"
    /codon_start=1
    /evidence=experimental
    /transl_table=11
    /product="nucleoside triphosphate transport protein 1"
    /protein_id="CA839534.1"
    /db_xref="GI:4538860"
    /translation="MTOTAEKPKFKLNSFLPPIHMLKYLPMFLMFCISNYTTL
    RDKDTLIVTAPGSGAEAIPIKILMLVPSAVVMLIYAKLSNLTNKOALFFAVLSPF
```

```
VFPAFPVIVYPCRHLHPFAADTLQSLSPGFMFIAMLRNMTFAVYVLSLWG
SVMLSLMFWGFANEITKISEAKRFALFGVANYALLISGPAITWSSKLRASLGEVD
PMGSLVPLMAMFLCSCAIIAACVWMMNRVYLTDPREFYPAELAKSKPKPMGSEF
SVILRSPYMLLALLIVICGICINLYETVMKSOIKMOPRNDVSAFPMGNFPMYGVY
SVFVWLFVGVNLTSTKRYALPDATKELAYIPLOEQCKKKAIDVVAAREFGSGS
MLAVVGAIQNLTSTKRYALPDATKELAYIPLOEQCKKKAIDVVAAREFGSGS
LIQGLLVGCSIGAMPFLVAFALIMWLTSAITKLNLKFLAASAKEDQELAEAA
AEKSSSAKESAPAEIGVS"
BASE COUNT 384 a 357 c 330 g 516 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 8.4e-10; Length 1587;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 AAGAAGTCTCGCAATGTTCTTAATGTTCTCTGATT 161
DB 73 AAGAAGTCTCGCAATGTTCTTAATGTTCTCTGATT 111

RESULT 6
AE001281/c 14168 bp DNA linear BCT 30-OCT-2000
LOCUS Chlamydia trachomatis section 8 of 87 of the complete genome.
DEFINITION AE001281 AE001273
ACCESSION AE001281
VERSION AE001281.1 GI:3328454
KEYWORDS
SOURCE
    Chlamydia trachomatis.
    Chlamydia trachomatis.
    Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 14168)
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis
JOURNAL Science 282 (5389), 754-759 (1998)
MEDLINE 9900809
PUBMED 9784136
TITLE 2 (bases 1 to 14168)
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
Comparative genomes of Chlamydia pneumoniae and C. trachomatis
Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 9920606
PUBMED 10192388
TITLE 3 (bases 1 to 14168)
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
Koonin,E.V. and Davis,R.W.
Direct Submission
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES
source
    /organism="Chlamydia trachomatis"
    /strain="D/UW-3/CX"
    /db_xref="taxon:813"
    /gene="flaA"
    /gene="flaA"
    /gene="flaA"
    /note="CT061"
    /note="86.847"
    /codon_start=1
    /transl_table=11
    /product="Sigma-28/whig Family"
    /protein_id="AAC67652.1"
    /db_xref="GI:3328455"
    /translation="MKTHDADIMQLYSTKRIHHRVLTSLYPLVKNVAHRLASGM
    PSYHKMEDIALCIVEGLIRAVERRDPKSKRPSYALFIITKATITGLRKQMDPPRSY
    YORANRLADMDISRQTLKEPDLGELCTELNISQDELSHWFSSSRPALVLSLHDFS
    CQDDDEGLAEERADRAETGYDVIRKKAISILITVALLADEKERQYVWALYYDDL
    VLKEIGKILGVSESRVSOIHSKALKLIRKLTLSLL"
```



```

Query Match          2.4%; Score 39; DB 1; Length 14168;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 AAGAAAGTTCTGCCAATGTTCTTAATGTTCTGTGATT 161
      |||
Db 7265 AAGAAAGTTCTGCCAATGTTCTTAATGTTCTGTGATT 7227

/product="Integral Membrane Protein"
/protein_id="AAC67661.1"
/db_xref="GI:3328464"

RESULT 7
LOCUS AX281684 5032 bp DNA linear PAT 02-NOV-2001
DEFINITION Sequence 93 from Patent WO0177389.
ACCESSION AX281684
VERSION AX281684.1 GI:16608935
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 Shiffman,D., Somogyi,R., Lawn,R., Sellhammer,J.J., Porter,G.J.,
TITLE Mikita,T. and Tai,J.
JOURNAL Genes expressed in foam cell differentiation
Patent: WO 0177389-A 93 18-OCT-2001;
FEATURES Incyte Genomics, Inc. (US)
SOURCE Location/Qualifiers
1. 5032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 445186.7"

BASE COUNT 1473 a 1026 c 1055 g 1477 t 1 others
ORIGIN

Query Match          1.3%; Score 21; DB 6; Length 5032;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 700 TAGATCCTTGGGGAATTCTT 720
      |||
Db 2927 TAGATCCTTGGGGAATTCTT 2947

RESULT 8
LOCUS AR064709 8220 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 11 from patent US 5849306.
ACCESSION AR064709
VERSION AR064709.1 GI:5994925
KEYWORDS
SOURCE unknown.
ORGANISM unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 8220)
TITLE Sim,K.Lee., Chichnis,C., Miller,L.H., Peterson,D.S., Su,X.-Z. and
JOURNAL Wellens,T.E.
FEATURES Binding domains from Plasmodium vivax and Plasmodium falciparum
erythrocyte binding proteins
Patent: US 5849306-A 11 15-DEC-1998;
SOURCE Location/Qualifiers
1. 8220
/organism="unknown"

BASE COUNT 3194 a 1249 c 1777 g 2000 t
ORIGIN

Query Match          1.3%; Score 21; DB 6; Length 8220;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 64 AAGAAACCTTTGGAAAT 84
      |||
Db 5046 AAGAAACCTTTGGAAAT 5066

RESULT 9
LOCUS AR089156 8220 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 11 from patent US 5993827.
ACCESSION AR089156
VERSION AR089156.1 GI:10015913
KEYWORDS
SOURCE unknown.
ORGANISM unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 8220)
TITLE Sim,K.Lee., Chichnis,C., Miller,L.H., Peterson,D.S., Su,X.-Z. and
JOURNAL Wellens,T.E.
FEATURES Binding domains from Plasmodium vivax and Plasmodium falciparum
erythrocyte binding proteins
Patent: US 5993827-A 11 30-NOV-1999;
SOURCE Location/Qualifiers
1. 8220
/organism="unknown"

BASE COUNT 3194 a 1249 c 1777 g 2000 t
ORIGIN

Query Match          1.3%; Score 21; DB 6; Length 8220;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 AAGAAACCTTTGGAAAT 84
      |||
Db 5046 AAGAAACCTTTGGAAAT 5066

RESULT 10
LOCUS AL592487 17462 bp DNA linear PRI 15-NOV-2001
DEFINITION Human DNA sequence from clone Rpl1-290b1 on chromosome 1, complete
sequence.
ACCESSION AL592487
VERSION AL592487.10 GI:16973138
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 17462)
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:16416250.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence

```

was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chrl>  
Rpl1-290B1 is from the library RPl1-11.2 constructed by the group of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rpl1-290B1 it may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone Rpl1-290B1 is at 1 in this sequence. The true left end of clone Rpl1-474021 is at 15463 in this sequence.

## FEATURES

source  
1..17462  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="Rpl1-290B1"  
/clone\_lib="RPl1-11.2"  
BASE COUNT 4402 a 4424 c 4149 g 4487 t  
ORIGIN

Query Match 1.3%; Score 21; DB 9; Length 17462;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 353 CTTTAAATTTCTTGCCCT 373  
|||||  
Db 15493 CTTTAAATTTCTTGCCCT 15473

RESULT 11  
PEAVARIA LOCUS 19124 bp DNA linear INV 14-SEP-1995  
DEFINITION Plasmodium falciiparum (strain Dd2) variant-specific surface protein (var-1) gene, complete cds.  
ACCESSION LA0608.1 GI:866374  
VERSION variant-specific surface protein.  
KEYWORDS Plasmodium falciiparum (strain Dd2) blood stage DNA.  
SOURCE variant-specific surface protein.  
ORGANISM Plasmodium falciiparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 19124)  
Su,X.Z., Heatwole,V.M., Wertheimer,S.P., Guinet,F., Herfeldt,J.A., Peterson,D.S., Ravetch,J.A. and Wellems,T.E.  
The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciiparum-infected erythrocytes  
Cell 82 (1), 89-100 (1995)  
JOURNAL MEDIAN 95330813  
PUBMED 7606788  
FEATURES

Location/Qualifiers  
1..19124  
/organism="Plasmodium falciiparum"  
/strain="Dd2"  
/db\_xref="taxon:5833"  
/def\_stage="blood stage"  
2068..2967  
/note="putative"  
/label="rif\_homology  
7307..17609  
/gene="var-1"  
join((7307..15139,16206..17609)  
/gene="var-1"  
/note="blood stage; pFEMP1 gene related; putative"  
/codon\_start=1  
/product="variant-specific surface protein"  
/protein\_id="AA7536.1"  
/db\_xref="GI:866375"  
/translation="MWELAKMGPEKAAGDDIEDSAKHMFDRIKGVYDKVEEAK  
RGKIGRLSEAKFEKNESQPTDECDLHKYHTVNTVINPCADRSDFRSDEY  
GGCCTHNRKIDSDGDNKACAPYRRLHVDONLEOIEPIKITNTNHLNLDVCAAKF

BASE COUNT 7824 a 2190 c 2790 g 6320 t  
ORIGIN

Query Match 1.3%; Score 21; DB 3; Length 19124;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 64 AAGAAAACCTTTGGAAT 84  
|||||  
Db 12366 AAGAAAACCTTTGGAAT 12386

RESULT 12  
AR089157

EGSITODYPKYATYGDSPSICITMLARSPADIGD1YVRGRLYLGNPOETIKOROLE  
MNTKTFGIYERKLNAGARNGNDPEPFKLBEDWMTNREYKATITCMKMGNTYFPA  
TNGNGERTKCYCRNDNOVPTPEYVYQYLMEFEWAEDEPRKKNKIKYGRNRGRK  
DKEDKDRICSNIGDCEKTRAIKLRGKQJISCLACNPEYVMINNOEBOPOKOK  
KYDEIKRYENGASGSRUKRDAGGTTTNDGEEKFEYBELNLSERYRDKFLEKLS  
NEEICTYVKEDEGGTIDFKNVNSDSTGASTANESGCTEYFRSKYCCPYCGVAKYN  
NGSSNMEERKNNKCKSGKLYEPKPEKGTITILSGGHDDIEERKLCDEKNG  
DTINGSGSGTGGSGNSGROLEYEERKCYKGEVDYVGHDEDEEDYEVANKNAGC  
ILKNOKNKEGGGNTSKREDEIOTKNTNPTTYVYAHMLKDSIMKKKLCORCLONGR  
ITCGNNCNDNCEBFKRWITOKRDEWKIVOHFTQNTIKGGSDNTEALPEPHDYV  
LOYNLOEFLKGSDEDAESEKENSLSAEAEELKHLREIIESDNNNOEASVGGVE  
OKNIMLNLNEXDEADLCLEIHEDEEEKKGGNGCIEEGENRPNPSCGESGNR  
YPLANKVAYOMHHKAKTOLASRAGRALRNDIOLAFONRGNSTJLKQJCKINENY  
SNDGRNSGGPCTGDDHGGTWRMRITKGGOTSGYKKNVFLPRRHHMTSL  
ENLDVGYTRKDKASHSILGDVQLAATDMAEITIKRKDONIDLTPTIOTKODENAK  
RAVRYSFADLGD11LRGDMWDEKSSYDMERLITVRNRIKEKDGJIKDNKYTGDS  
KAPAYKRLRADWMEANRHOVWRAMKCATKIGICGMVDDVITPQRLKMNEMAWMYK  
AOSOEYDKLKKIKADCMKSGDKCTGQDVDCGKCAACDKYKEIEERKMNQMRISDK  
YNLYLOAKTSTNPRGVLGDDDPDYQOVMYDFTPLIHKASIAARVLYKRAASPTET  
AAAPITPYSYTAAGYTHOETCYGCCOCTOPECKKHGATSTSTKREKTYRQPPPE  
YATACDCINSQTEEPKKKEENVESACKIYKILLEGKNGRTVEECNPKESYEDMDCK  
NINIDISHGACMPERRKOLCLYVIAHESQTEINITDNLDAEITKTAAEFTLSYOT  
KSKNDSEAKILDRGLIPSOFLRSMVYTFGDRDCLNTDLSKKNQDAKADKIGKFE  
SKDSKSPSGLSROEMKKTNGPEIMKGCALATKYVDTDNRKIKNDYSDKYNOQ  
NGNPSLEEFKAPQFLRMIMIEGGEFCAEROKKENIYKDACNEINSGOCDADAKHRC  
OACRAYOYEVENKREKFEFGOTNNFVLKANOPODEYKGYKGYKQVPIQCNEXYLQK  
CDNNKSCMDGNVLSVSEKPEFKYAHKYPEKDCYQGGKHYPSITPPPEYVQPEA  
PTVTVDVCSIVKTLKDTNNFSDACGLKYGTAPSMMKCIPTSDRSGAGATGKSGSD  
SGSICIPPRRRRLYVGLQEWATLAPGEGAPSHSRDALRNAPFIOAALETFFLMD  
RYKEERKPOGDSOOLASOLSTSYSDDEEDPDRLONGKLPDFLRIMEFTLGDYRD  
ILVHGNTSDSGNTGNSNNNIVLEASGNKEDMOKIOEKIOLILPKNGGPLVYSSA  
OTPRKMNNEHAEISTKMGICALATYTERKNPTDASADENKIEKDEYVEKPEGRADH  
GTATPTGTGYTQYDIEKYLEDTSGAKTPASSDTLDSFVLRPPYFRILEMGN  
FCRRKRLKLOIRHECVVEENGGSRRGITTROYSGGEGACNEMLPKNDGVPLERK  
SCAPKCSYRKMIESKGEFEKOEKAYEODKCVNOSNKHDFCETLITSSNAKOF  
LKTGPKCPNVNEKGTIFDDDKTRFKHNRDPCPLKFCYVNCIEGIDRDEMGRVCGY  
VYCKPEVNGYKGGKHIIOLIRALYRWYVEFEYENKIKHISRINGETISPCIKK  
VEKWVDOKRKEKERTERKQOYKNDSDDNVSFLETLIPOTTDANAKRYIKLSC  
FGNSGCSASANDONKNGEYKDAIDCMKKLKDITGEEKHNHDTSDIESQTPOT  
LEDETLLDDILETEBAKNNMPKICEVNLKTAQODEGCVPAENSEEPAVDSKETP  
ROPTVLPKEEVAVEPPPPPOEKAAPAPIPQPOPPTPTOLDBPHVLTALVSTLAW  
SVGIGFATPTFYFLKTKSSVGNLFIOLIPKSDDYDIPKLSNRYIPTYSGYRGR  
RYTILEGSGTDSGYTDHYSDITSSSEYEMDINDIYVDSPPYKTLIEVLEPSGN  
NFTVASGNNTASGNNTASGNTPSPDNDONONGTISSKTTDENNOJLDEFTSOL  
OSEPNTERNMLGTVNDNTHPTTSHHVAEKFPTMSIHDRNLSEGEYNTMDPSGNP  
PINISDSTNSMDSLTSNNHSPYNDKNLYSGIDLINDALSGNHLIDIDEMLRKNEEL  
FGTGHNRKHTNTYVWAKPARDDPTTNOINLPHKWLDRHDMCEKMKNNHETPLKEL  
WENETHSGDINSIGIPSGNHVLTDSOIODMDNKTKEINTEITNMDKSTMDTILDD  
LEKYNPEYVDFYEDDILYHDVDEKSMOYIYDHNNTNNMNDVPTKMHIEKNITV  
NKKETFEFEYPISDIWNIT"

exon  
/gene="var-1"  
/number=1  
15140..16205  
/gene="var-1"  
16206..>17609  
/gene="var-1"  
/number=2

intron  
/gene="var-1"  
16206..>17609  
/gene="var-1"

exon





```

* 23207 23638: contig of 432 bp in length
*      gap of unknown length
* 23639 24265: contig of 627 bp in length
*      gap of unknown length
* 24266 25200: contig of 935 bp in length
*      gap of unknown length
* 25201 26153: contig of 953 bp in length
*      gap of unknown length
* 26154 26999: contig of 846 bp in length
*      gap of unknown length
* 27000 27419: contig of 420 bp in length
*      gap of unknown length
* 27420 28185: contig of 766 bp in length
*      gap of unknown length
* 28186 29128: contig of 943 bp in length
*      gap of unknown length
* 29129 29817: contig of 689 bp in length
*      gap of unknown length
* 29818 30211: contig of 394 bp in length
*      gap of unknown length
* 30212 30820: contig of 609 bp in length
*      gap of unknown length
* 30821 31969: contig of 1149 bp in length
*      gap of unknown length
* 31970 32714: contig of 745 bp in length
*      gap of unknown length
* 32715 34020: contig of 1306 bp in length
*      gap of unknown length
* 34021 35177: contig of 1157 bp in length
*      gap of unknown length
* 35178 36602: contig of 1425 bp in length
*      gap of unknown length
* 36603 37916: contig of 1314 bp in length
*      gap of unknown length
* 37917 38222: contig of 306 bp in length
*      gap of unknown length
* 38223 39014: contig of 792 bp in length
*      gap of unknown length
* 39015 40181: contig of 1167 bp in length
*      gap of unknown length
* 40182 41359: contig of 1178 bp in length
*      gap of unknown length
* 41360 42186: contig of 827 bp in length
*      gap of unknown length
* 42187 42987: contig of 801 bp in length
*      gap of unknown length
* 42988 44029: contig of 1042 bp in length
*      gap of unknown length
* 44030 45441: contig of 1412 bp in length
*      gap of unknown length
* 45442 46864: contig of 1423 bp in length
*      gap of unknown length
* 46865 48255: contig of 1391 bp in length
*      gap of unknown length
* 48256 49365: contig of 1110 bp in length
*      gap of unknown length
* 49366 50340: contig of 975 bp in length
*      gap of unknown length
* 50341 51309: contig of 969 bp in length
*      gap of unknown length
* 51310 52722: contig of 1413 bp in length
*      gap of unknown length
* 52723 54001: contig of 1279 bp in length
*      gap of unknown length
* 54002 55737: contig of 1736 bp in length
*      gap of unknown length
* 55738 57052: contig of 1315 bp in length
*      gap of unknown length
* 57053 58809: contig of 1757 bp in length
*      gap of unknown length
* 58810 60365: contig of 1556 bp in length
*      gap of unknown length
* 60366 62096: contig of 1731 bp in length

```

```

*      62097 64348: gap of unknown length
*      64349 65921: gap of unknown length
*      65922 67231: gap of unknown length
*      67232 69525: gap of unknown length
*      69526 70784: gap of unknown length
*      70785 73256: contig of 2472 bp in length

Query Match      1.3%  Score 21; DB 2; Length 90106;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATCAAAAGACACTTATTG 207
Db 45690 ATCAAAAGACACTTATTG 45670
|||||
RESULT 14
AC092392 98286 bp DNA linear PRI 15-DEC-2001
LOCUS Homo sapiens chromosome 1 clone RP4-544H6, complete sequence.
AC092392 AL109625
AC092392.2 GI:17861066
DEFINITION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 98286)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and
Haugen, E.D.
Direct Submision
Submitted (03-JUL-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 98286)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C. and
Haugen, E.D.
Direct Submision
Submitted (15-DEC-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Dec 15, 2001 this sequence version replaced gi:14589610.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UMGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP4-544H6 (sc0188)
----- Summary Statistics
Sequencing vector: plasmid; 22% of reads
Sequencing vector: plasmid; 78% of reads
Chemistry: Dye-terminator ET; 32% of reads
Chemistry: Dye-terminator Big Dye; 68% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 98273 bases at least Q40
Consensus quality: 98278 bases at least Q30
Consensus quality: 98286 bases at least Q20
Insert size: 98287; sum-of-contigs
Quality coverage: 9.1x in Q20 bases; sum-of-contigs
-----
Overlapping Sequences:

```

5' : RP5-1043L3 (UMGC:sc0545) AC099062  
3' : RP4-671122 (UMGC:sc0205) AL356298

# Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:  
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector. In order to accurately represent the entire circular BAC, small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII

EcoRI

HindIII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----
6023	5830	8696	8978	1849	1934
-----	-----	-----	-----	-----	-----
2067	2146	6101	6394	6382	6210
-----	-----	-----	-----	-----	-----
7084	7102	6972	6986	512	<800
-----	-----	-----	-----	-----	-----
2144	2146	6986	6986	449	<800
-----	-----	-----	-----	-----	-----
1462	1431	11036	10738	1792	1934
-----	-----	-----	-----	-----	-----
2877	2772	826	827	7699	7751
-----	-----	-----	-----	-----	-----
252	<800	1438	1380	3248	3382
-----	-----	-----	-----	-----	-----
547	<800	2212	2194	6225	6210
-----	-----	-----	-----	-----	-----
815	883	1590	1521	9	<800
-----	-----	-----	-----	-----	-----
308	<800	3225	3154	720	726
-----	-----	-----	-----	-----	-----
515	<800	3261	3312	1705	1735
-----	-----	-----	-----	-----	-----
1321	1320	671	<800	2231	2162
-----	-----	-----	-----	-----	-----
11034	10897	2171	2194	83	<800
-----	-----	-----	-----	-----	-----
4803	4847	3889	3911	298	<800
-----	-----	-----	-----	-----	-----
1643	1622	2401	2382	307	<800
-----	-----	-----	-----	-----	-----
2140	2146	742	827	3417	3382
-----	-----	-----	-----	-----	-----
2925	2947	8122	8089	2700	2801
-----	-----	-----	-----	-----	-----
4658	4573	988	974	262	<800

-----	3569	3631	2263	2194	1407	1374
-----	3936	4009	825	827	5580	5488
-----	653	<800	7300	7287	2031	2162
-----	18	<800	1900	1898	2574	2646
-----	1053	1111	4808	4762	2085	2162
-----	1135	1111	2669	2653	897	907
-----	50	<800	2712	2746	202	<800
-----	113	<800	6452	6394	706	<800
-----	2617	2615	2160	2194	3060	3207
-----	689	<800	3777	3706	3662	3618
-----	402	<800	792	<800	1702	1655
-----	7554	7725	-----	-----	250	<800
-----	592	<800	-----	-----	1010	1005
-----	5487	5529	-----	-----	333	<800
-----	1643	1622	-----	-----	233	<800
-----	6052	6239	-----	-----	1223	1200
-----	896	883	-----	-----	4551	4630
-----	5326	5322	-----	-----	1948	1934
-----	1036	1030	-----	-----	4120	3964
-----	6826	6811	-----	-----	7902	8074
-----	1378	1320	-----	-----	5161	5086
-----	599	<800	-----	-----	543	<800
-----	2743	2615	-----	-----	6163	6210
-----	-----	-----	-----	-----	5260	5086
-----	-----	-----	-----	-----	1692	1655
-----	-----	-----	-----	-----	2232	2162
-----	-----	-----	-----	-----	451	<800
-----	-----	-----	-----	-----	119	<800

## FEATURES

source

Location/Qualifiers  
1. 98286  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP4-544H6"  
/clone\_lib="RPCI human PAC library 4"

BASE COUNT 27438 a 19511 c 20578 g 30759 t  
ORIGIN

Query Match 1.3%; Score 21; DB 9; Length 98286;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 TAGATCCTTGGGAGATTTCCTT 720  
|||||

Db 60168 TAGATCCTTGCGGATTCTT 60188

RESULT 15  
AC113803

LOCUS AC113803 126996 bp DNA linear HTG 17-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-293f17, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 52 unordered pieces.

AC113803  
AC113803.3 GI:21744586  
HTG: HTGS\_PHASE1.

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Rattus norvegicus  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 126996)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,K., Banks,T.,  
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Burch,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,  
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlssoon,E., Kelly,S., Khan,U., King,U., Korvah,J., Kovar,C.,  
Krtovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Lousaged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,N., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogun,M., Okunnu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rivas,M., Rojas,A., RojuboKan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherrer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,  
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Umsani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 126996)  
Direct Submission  
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 13, 2002 this sequence version replaced gi:19525967.  
Genome Center

Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Center project name: GSHO  
Center clone name: CH230-293f17

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329

Consensus quality: 72236 bases at least Q40  
Consensus quality: 75069 bases at least Q30  
Consensus quality: 77416 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently  
consists of 52 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1	1036:	contig of 1036 bp in length
1037	1136:	gap of unknown length
1137	2213:	contig of 1077 bp in length
2214	2313:	gap of unknown length
3914	3997:	contig of 1684 bp in length
3998	4097:	gap of unknown length
4098	5272:	contig of 1175 bp in length
5273	5372:	gap of unknown length
5373	6575:	contig of 1203 bp in length
6576	6675:	gap of unknown length
6676	8156:	contig of 1481 bp in length
8157	8256:	gap of unknown length
8257	9450:	contig of 1194 bp in length
9451	9550:	gap of unknown length
9551	10960:	contig of 1410 bp in length
10961	11060:	gap of unknown length
11061	12118:	contig of 1058 bp in length
12119	12218:	gap of unknown length
12219	13878:	contig of 1660 bp in length
13879	13978:	gap of unknown length
13979	15252:	contig of 1274 bp in length
15253	15352:	gap of unknown length
15353	16951:	contig of 1599 bp in length
16952	17051:	gap of unknown length
17052	18138:	contig of 1087 bp in length
18139	18238:	gap of unknown length
18239	19414:	contig of 1176 bp in length
19415	19514:	gap of unknown length
19515	21376:	contig of 1862 bp in length
21377	21476:	gap of unknown length
21477	23039:	contig of 1563 bp in length
23040	23139:	gap of unknown length
23140	25010:	contig of 1871 bp in length
25011	25110:	gap of unknown length
25111	26280:	contig of 1170 bp in length
26281	26380:	gap of unknown length
26381	27993:	contig of 1613 bp in length
27994	28093:	gap of unknown length
28094	30038:	contig of 1945 bp in length
30039	30138:	gap of unknown length
30139	31754:	contig of 1616 bp in length
31755	31854:	gap of unknown length
31855	33816:	contig of 1962 bp in length
33817	33916:	gap of unknown length
33917	36417:	contig of 2501 bp in length
36418	36517:	gap of unknown length
36518	38336:	contig of 1819 bp in length
38337	38436:	gap of unknown length

\* 38437 39646: contig of 1210 bp in length  
\* 39647 39746: gap of unknown length  
\* 39747 41954: contig of 2208 bp in length  
\* 41955 42054: gap of unknown length  
\* 42055 44188: contig of 2134 bp in length  
\* 44189 44288: gap of unknown length  
\* 44289 47196: contig of 2908 bp in length  
\* 47197 47296: gap of unknown length  
\* 47297 49882: contig of 2586 bp in length  
\* 49883 49982: gap of unknown length  
\* 49983 51747: contig of 1765 bp in length  
\* 51748 51847: gap of unknown length  
\* 51848 54070: contig of 2223 bp in length  
\* 54071 54170: gap of unknown length  
\* 54171 55692: contig of 1522 bp in length  
\* 55693 55792: gap of unknown length  
\* 55793 58784: contig of 2992 bp in length  
\* 58785 58884: gap of unknown length  
\* 58885 60823: contig of 1939 bp in length  
\* 60824 60923: gap of unknown length  
\* 60924 62885: contig of 1962 bp in length  
\* 62886 62985: gap of unknown length  
\* 62986 64984: contig of 1999 bp in length  
\* 64985 65084: gap of unknown length  
\* 65085 67271: contig of 2187 bp in length  
\* 67272 67371: gap of unknown length  
\* 67372 70966: contig of 3595 bp in length  
\* 70967 71066: gap of unknown length  
\* 71067 74043: contig of 2977 bp in length  
\* 74044 74143: gap of unknown length  
\* 74144 77759: contig of 3616 bp in length  
\* 77760 77859: gap of unknown length  
\* 77860 79868: contig of 2009 bp in length  
\* 79869 79968: gap of unknown length  
\* 79969 83800: contig of 3832 bp in length  
\* 83801 83900: gap of unknown length  
\* 83901 86500: contig of 2600 bp in length  
\* 86501 86600: gap of unknown length  
\* 86601 89461: contig of 2861 bp in length  
\* 89462 89561: gap of unknown length  
\* 89562 93647: contig of 4086 bp in length  
\* 93648 93747: gap of unknown length  
\* 93748 97544: contig of 3797 bp in length  
\* 97545 97644: gap of unknown length  
\* 97645 101973: contig of 4329 bp in length  
\* 101974 102073: gap of unknown length  
\* 102074 106111: contig of 4038 bp in length  
\* 106112 106211: gap of unknown length  
\* 106212 112152: contig of 5941 bp in length  
\* 112153 112252: gap of unknown length  
\* 112253 116920: contig of 4668 bp in length  
\* 116921 117020: gap of unknown length  
\* 120957: contig of 3937 bp in length  
\* 120958 121057: gap of unknown length  
\* 121058 126996: contig of 5939 bp in length.

FEATURES  
source  
1. .126996  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"

Query Match 1.3%, Score 21; DB 2; Length 126996;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1576 ATTCAGCTCTGCTTCTTCAT 1596  
|||||

Db 43370 ATTCAGCTCTGCTTCTTCAT 43390

RESULT 16  
AC007335/c AC007335 134482 bp DNA linear PRI 03-NOV-2000  
LOCUS Homo sapiens chromosome 16 clone RP11-165M2, complete sequence.  
DEFINITION

ACCESSION AC007335  
VERSION AC007335.6 GI:11079401  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
2 (bases 1 to 134482)  
Brune,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,  
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,  
McCurry,K., Han,C. and Deaven,L.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
3 (bases 1 to 134482)  
Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA

REFERENCE  
AUTHORS  
DOE Joint Genome Institute.

TITLE  
JOURNAL  
COMMENT  
Submitted (03-NOV-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Nov 3, 2000 this sequence version replaced gi:9795550.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

FEATURES  
source  
1. .134482  
/organism="Homo sapiens"  
/db\_xref="taxon:3606"  
/chromosome="16"  
/clone="RP11-165M2"

BASE COUNT 35905 a 28147 c 29995 g 40435 t  
ORIGIN

Query Match 1.3%, Score 21; DB 9; Length 134482;  
Best Local Similarity 100.0%; Pred. No. 9.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 187 ATCAAAAGACACTTATTG 207  
|||||

Db 78708 ATCAAAAGACACTTATTG 78688

RESULT 17  
AC104838 143279 bp DNA linear PRI 31-MAY-2002  
LOCUS Homo sapiens chromosome 1 clone RP4-671I22, complete sequence.  
AC104838 AL356238  
ACCESSION AC104838.2 GI:21281548  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM human.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
1 (bases 1 to 143279)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
2 (bases 1 to 143279)  
Submitted (21-DEC-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
3 (bases 1 to 143279)  
Submitted (21-DEC-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z., Saepthimachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C. and Haugen, E.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-MAY-2002) Genome Center, University of Washington, Box 353145, Seattle, WA 98195, USA  
 COMMENT On May 31, 2002 this sequence version replaced gl:17976476.  
 ----- Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UMG  
 Web site: <http://www.genome.washington.edu>  
 Contact: [uwgchits@u.washington.edu](mailto:uwgchits@u.washington.edu)  
 Drafting Center: SC  
 ----- Project Information  
 Center project name: chr-1  
 Center clone name: RP4-671122 (sc0205)  
 ----- Summary Statistics  
 Sequencing vector: plasmid; 70% of reads  
 Sequencing vector: plasmid; 108752; 30% of reads  
 Chemistry: Dye-terminator ET; 59% of reads  
 Chemistry: Dye-terminator Big Dye; 41% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 143182 bases at least Q40  
 Consensus quality: 143276 bases at least Q30  
 Consensus quality: 143278 bases at least Q20  
 Insert size: 143279; sum-of-contigs  
 Quality coverage: 9.3x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:  
 5': Mapping in progress  
 3': RP11-442N6 (UMGC:sc0307) AC103552, 43012-bp overlap

----- Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:  
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

----- Ecoli

----- HindIII

----- BglII

----- SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

----- 9950 10064 1637 1634 8109 8153  
 ----- 2184 2191 449 <800 5671 5728  
 ----- 8065 8142 512 <800 8094 8153

486	<800	2814	2847	592	<800
106	<800	1247	1206	5487	5342
7113	7150	4695	4723	1643	1649
988	992	3373	3403	7390	7505
2263	2191	6390	6351	896	888
825	842	1948	1923	5327	5342
8638	8681	4120	4010	1036	1094
1900	1918	9240	9293	6826	6849
4809	4785	5162	5133	1378	1354
2669	2648	543	<800	599	<800
2712	2773	6163	6351	2743	2754
6452	6486	5260	5133	3995	3978
2160	2191	1692	1634	1146	1094
3777	3727	2232	2178	7	<800
4526	4454	451	<800	1613	1649
4195	4163	119	<800	3589	3609
117	<800	17870	17825	6332	6421
273	<800	440	<800	6781	6849
3560	3601	844	843	215	<800
5446	5428	670	<800	6	<800
6564	6486	655	<800	1740	1649
697	<800	849	843	3656	3609
10283	10064	4779	4723	1657	1649
720	<800	343	<800	1875	1868
4163	4163	779	763	4873	4794
1386	1434	6371	6351	826	823
1472	1434	1554	1513	720	728
1214	1182	221	<800	321	<800
678	<800	4724	4723	69	<800
50	<800	1220	1206	1820	1785
3978	3963	3375	3403	3084	3148
6193	6203	384	<800	1525	1492
76	<800	5108	5133	4613	4568
2441	2474	686	<800	818	823
1446	1434	764	763	3366	3409
2575	2648	13713	13478	7051	7129
4884	4914	3922	4010	2376	2386

```

-----
3579      3601      11468      11161      -----
-----
2672      2648      20395      20637      4222      4259
-----
4010      3963      753      763      2338      2386
-----
3229      3293      687      <800      2157      2174
-----
2172      2191      -----
-----
6136      6203      -----
-----
1284      1266      -----
-----
1993      2009      -----
-----
3412      3396      -----
-----

```

**FEATURES**  
**source** 1.143279

**Query Match** 1.3%; Score 21; DB 9; Length 143279;  
**Best Local Similarity** 100.0%; Pred. No. 9.9;  
**Matches** 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Qy** 700 TAGATCCTTGGGGAATTCTT 720  
 |||||||  
**Db** 3467 TAGATCCTTGGGGAATTCTT 3487

**RESULT 18**  
**AC046148** 150495 bp DNA linear HTG 16-OCT-2001  
**LOCUS** Mus musculus chromosome 14 clone RP23-265D21, \*\*\* SEQUENCING IN  
**DEFINITION** PROGRESS \*\*\*, 33 unordered pieces.  
**ACCESSION** AC046148 GI:16118090  
**VERSION** AC046148.6  
**KEYWORDS** HTG; HTGS\_PHASE1.  
**SOURCE** Mus musculus.  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 150495)  
 Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,  
 Dedert, D., Thomas, S., Okwuonu, G., Carllock, C., Garner, T.,  
 Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,  
 Bunay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,  
 Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,  
 Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,  
 Gottrell, J., H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,  
 Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,  
 Kovar, C., Liu, J., Liu, W., Louised, H., Lozano, R., Martin, R.,  
 Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,  
 Neel, D., Nelson, A., Nguyen, R., Nguyen, N., Oguri, M., Parish, B.,  
 Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, J., Wallington, S.,  
 Williamson, A., Wrenford, G., Zhou, X., Bouck, J., Hodgson, A.,  
 Munzy, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,  
 Morley, K. and Gibbs, R.  
 Direct Submission  
 Unpublished

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**

**COMMENT**

2 (bases 1 to 150495)  
 Worley, K.C.  
 Direct Submission  
 Submitted (13-APR-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Oct 14, 2001 this sequence version replaced gi:11094639.  
 -----  
 Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 -----  
 Project Information  
 Center project name: MAEB  
 Center clone name: RP23-265D21  
 -----  
 Summary Statistics  
 Sequencing vector: M13, 108821  
 Chemistry: Dye-Primer Bodipy: 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 132230 bases at least Q40  
 Consensus quality: 149333 bases at least Q30  
 Consensus quality: 156096 bases at least Q20  
 Estimated insert size: 153840; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 33 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 10853: contig of 10853 bp in length  
 10854  
 10953: gap of unknown length  
 10954  
 20346: contig of 9393 bp in length  
 20347  
 20446: gap of unknown length  
 20447  
 30391: contig of 9945 bp in length  
 30392  
 30491: gap of unknown length  
 30492  
 39349: contig of 8858 bp in length  
 39350  
 39449: gap of unknown length  
 39450  
 47183: contig of 7734 bp in length  
 47184  
 47283: gap of unknown length  
 47284  
 53762: contig of 6479 bp in length  
 53763  
 53862: gap of unknown length  
 53863  
 58693: contig of 4807 bp in length  
 58694  
 58769: gap of unknown length  
 58770  
 65058: contig of 6289 bp in length  
 65059  
 65158: gap of unknown length  
 65159  
 69830: contig of 4672 bp in length  
 69831  
 69931: gap of unknown length  
 69932  
 73910: contig of 3980 bp in length  
 73911  
 74010: gap of unknown length  
 74011  
 78784: contig of 4674 bp in length  
 78785  
 83343: gap of unknown length  
 83344  
 83443: contig of 4559 bp in length  
 83444  
 83444: gap of unknown length  
 83445  
 87849: contig of 4406 bp in length  
 87850  
 87949: gap of unknown length  
 87950  
 92076: contig of 4127 bp in length  
 92077  
 92176: gap of unknown length  
 92177  
 97227: contig of 5051 bp in length  
 97228  
 97327: gap of unknown length  
 97328  
 101356: contig of 4029 bp in length  
 101357  
 101456: gap of unknown length  
 101457  
 104693: contig of 3237 bp in length  
 104694  
 104793: gap of unknown length  
 104794  
 108132: contig of 3339 bp in length  
 108133  
 108232: gap of unknown length  
 108233  
 111747: contig of 3515 bp in length

```

* 111748 111847: gap of unknown length
* 111848 115180: contig of 3333 bp in length
* 111849 115280: gap of unknown length
* 115281 118548: contig of 3268 bp in length
* 118549 118648: gap of unknown length
* 118649 121539: contig of 2891 bp in length
* 121540 121640: gap of unknown length
* 121640 124526: contig of 2887 bp in length
* 124527 124626: gap of unknown length
* 124627 127512: contig of 2886 bp in length
* 127513 127612: gap of unknown length
* 127613 131027: contig of 3415 bp in length
* 131028 131127: gap of unknown length
* 131128 133604: contig of 2477 bp in length
* 133605 133704: gap of unknown length
* 133705 135894: contig of 2190 bp in length
* 135895 135994: gap of unknown length
* 135995 138009: contig of 2015 bp in length
* 138010 138109: gap of unknown length
* 138110 140502: contig of 2393 bp in length
* 140503 140602: gap of unknown length
* 140603 142745: contig of 2143 bp in length
* 142746 142845: gap of unknown length
* 142846 145293: contig of 2448 bp in length
* 145294 145393: gap of unknown length
* 145394 148334: contig of 2941 bp in length
* 148335 148434: gap of unknown length
* 148435 150495: contig of 2061 bp in length.
Location/Qualifiers
1. 150495
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-265D21"

```

```

BASE COUNT 40913 a 32705 c 32685 g 40987 t 3205 others
ORIGIN

```

```

Query Match 1.3%: Score 21; DB 2; Length 150495;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1575 GATTCAGCTCTGCTCTTCA 1595
|||||
DB 129028 GATTCAGCTCTGCTCTTCA 129048

```

```

RESULT 19
AL513016 150720 bp DNA linear PRI 02-NOV-2001
LOCUS Human DNA sequence from clone Rp11-474021 on chromosome 1, complete
DEFINITION
ACCESSION AL513016
VERSION AL513016.5 GI:16605693
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 150720)
REFERENCE 1
AUTHORS Ramsay, H.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hinxton@sanger.ac.uk

```

On Nov 3, 2001 this sequence version replaced gi:16501152. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Mp., MORNPEP; Information on the MORNPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/BCP/chr1> Rp11-474021 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6

```

FEATURES
source
1. 150720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rp11-474021"
/chromosome="1"

```

```

misc_feature 47400..47444
/note="Sequence from overlapping clone Rp11-290B1 (AL592487). Assembly confirmed by restriction digest."

```

```

BASE COUNT 40472 a 33567 c 34757 g 41924 t
ORIGIN

```

```

Query Match 1.3%: Score 21; DB 9; Length 150720;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 353 CTTTAAATTTCTTGCCCT 373
|||||
DB 31 CTTTAAATTTCTTGCCCT 11

```

```

RESULT 20
AC101941 158599 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone Rp24-112F21, WORKING DRAFT SEQUENCE, 23
DEFINITION
ACCESSION AC101941
VERSION AC101941.2 GI:22381377
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 158599)
REFERENCE 1
AUTHORS Birren, B., Nussbaum, C., and Lander, E.
TITLE Unpublished
JOURNAL 2 (bases 1 to 158599)

```

Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barua, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collange, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DePellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Glinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Menus, L., Milhova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

**TITLE**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**

Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 158599)  
 Biren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Charato, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fardo, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Milhova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

**TITLE**  
**JOURNAL**  
**COMMENT**

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 21, 2002 this sequence version replaced g1:117060717.  
 All repeats were identified using RepeatMasker:  
 Smt, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center -----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIDR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information -----  
 Center project name: L17769  
 Center clone name: 112.F.21  
 ----- Summary Statistics -----  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 150624 bases at least Q40  
 Consensus quality: 154269 bases at least Q30  
 Consensus quality: 155666 bases at least Q20  
 Insert size: 147000; agarose-fp  
 Insert size: 156399; sum-of-ctrls  
 Quality coverage: 6.0 in Q20 bases; agarose-fp  
 Quality coverage: 5.7 in Q20 bases; sum-of-ctrls  
 ----- NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----  
 1 13026: contig of 13026 bp in length  
 \* 13027 13126: gap of 100 bp  
 \* 13127 13831: contig of 705 bp in length  
 \* 13832 13931: gap of 100 bp

13932 14601: contig of 670 bp in length  
 \* 14602 14701: gap of 100 bp  
 \* 14702 15451: contig of 750 bp in length  
 \* 15452 15551: gap of 100 bp  
 \* 15552 16678: contig of 1127 bp in length  
 \* 16679 16778: gap of 100 bp  
 \* 16779 17604: contig of 826 bp in length  
 \* 17605 17704: gap of 100 bp  
 \* 17705 18923: contig of 1219 bp in length  
 \* 18924 19023: gap of 100 bp  
 \* 19024 20310: contig of 1287 bp in length  
 \* 20311 20410: gap of 100 bp  
 \* 20411 22661: contig of 2251 bp in length  
 \* 22662 22761: gap of 100 bp  
 \* 22762 23949: contig of 1188 bp in length  
 \* 23950 24049: gap of 100 bp  
 \* 24050 50090: contig of 26041 bp in length  
 \* 50091 50190: gap of 100 bp  
 \* 50191 51831: contig of 1641 bp in length  
 \* 51832 51931: gap of 100 bp  
 \* 51932 55850: contig of 3919 bp in length  
 \* 55851 55950: gap of 100 bp  
 \* 55951 59837: contig of 3887 bp in length  
 \* 59838 59937: gap of 100 bp  
 \* 59938 63375: contig of 3438 bp in length  
 \* 63376 63475: gap of 100 bp  
 \* 63476 71452: contig of 7977 bp in length  
 \* 71453 71552: gap of 100 bp  
 \* 71553 80301: contig of 8749 bp in length  
 \* 80302 80401: gap of 100 bp  
 \* 80402 90305: contig of 9904 bp in length  
 \* 90306 90405: gap of 100 bp  
 \* 90406 100251: contig of 9846 bp in length  
 \* 100252 100351: gap of 100 bp  
 \* 100352 113788: contig of 13437 bp in length  
 \* 113789 113888: gap of 100 bp  
 \* 113889 134105: contig of 20217 bp in length  
 \* 134106 134205: gap of 100 bp  
 \* 134206 157288: contig of 23083 bp in length  
 \* 157289 157388: gap of 100 bp  
 \* 157389 158599: contig of 1211 bp in length.

**FEATURES**  
**SOURCE**  
 Location/Qualifiers  
 1..158599  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP24-112F21"  
 /clone\_lib="RPCI-24 Male Mouse BAC"  
 1..13026  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 13127..13831  
 /note="assembly\_fragment"  
 13932..14601  
 /note="assembly\_fragment"  
 14702..15451  
 /note="assembly\_fragment"  
 15552..16678  
 /note="assembly\_fragment"  
 16779..17604  
 /note="assembly\_fragment"  
 17705..18923  
 /note="assembly\_fragment"  
 19024..20310  
 /note="assembly\_fragment"  
 20411..22661  
 /note="assembly\_fragment"  
 22762..23949  
 /note="assembly\_fragment"  
 24050..50090  
 /note="assembly\_fragment"  
 50191..51831  
 /note="assembly\_fragment"



```

BASE COUNT      45140 a 34853 c 34643 g 41783 t 2200 others
ORIGIN
misc_feature    51932..55850
                /note="assembly_fragment"
misc_feature    55951..59837
                /note="assembly_fragment"
misc_feature    59938..63375
                /note="assembly_fragment"
misc_feature    63476..71452
                /note="assembly_fragment"
misc_feature    71553..80301
                /note="assembly_fragment"
misc_feature    80402..90305
                /note="assembly_fragment"
misc_feature    90406..100251
                /note="assembly_fragment"
misc_feature    100352..113788
                /note="assembly_fragment"
misc_feature    113889..134105
                /note="assembly_fragment"
misc_feature    134206..157288
                /note="assembly_fragment"
misc_feature    157389..158599
                /note="assembly_fragment"
                clone_end=T7
                vector_side=right

```

Query Match	1.3%	Score 21:	DB 2:	Length 158559:
Best Local Similarity	100.0%	Pred. No. 9.9:		
Matches 21:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:

QY 1436 TCGAGCTATTGGAGCTATGAC 1456  
|||||  
Db 104988 TCGAGCTATTGGAGCTATGAC 105008

RESULT	21
AC128538	
LOCUS	
DEFINITION	AC128538 165679 bp DNA linear HTG 19-JUL-200*
	Rattus norvegicus clone CHZ30-402010. *** SEQUENCING IN PROGRESS
***,	41 unordered pieces.

ACCESSION	AC128538
VERSION	AC128538.1
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Rattus norvegicus.
ORGANISM	Rattus norvegicus

REFERENCE  
AUTHORS

1 (bases 1 to 165679)  
Muzny D. M., Adams, C., Adio-Oduola, B., Ali-Usman, F. R., Allen, C., Alsbrooks, S. L., Amaralunga, H. C., Are, J. R., Ayele, M., Banks, T., Barbieri, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnah, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P., Buay, C., Burch, P., Buttel, C., Butrell, K. L., Byrd, N. C., Carrott, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Day-Carroll, L., Deedrich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinu, H., Douthwaite, K. J., Draper, H., Dugan, L., Edwards, S., Durbin, K. J., Earhart, C., Edgar, C., Edwards, C. C., Elhaj, C., Escoto, M., Falls, T., Ferragudo, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsti, F., Howard, S., Huber, J., Huily, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kralovic, J., Kireshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, Y., Li, Z., Lichteager, O., Lieu, C., Liu, J., Liu, W., Louisgeed, H., Lozada, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,

TITLE  
JOURNAL  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL

COMMENT

<p>                 Maheshwari,M., Mapua,P., Martin,R., Mattindale,A., Martinez,E.,                  Massey,E., Mawlini,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,                  Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,                  Nguyen,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,                  Nguyen,N., Nickerson,E., Nwokweto,S., Ogun,M., Okunolu,G.,                  Oragunye,N., Oyler,R., Pace,A., Payton,B., Peery,J., Perez,L.,                  Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,                  Rives,M., Rojas,A., Kojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,                  Scherer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,I.,                  Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,                  Sutton,A., Swakek,A., Tabori,P., Tamerisa,A., Tameris,K., Tang,H.,                  Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,                  Usmani,R., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q.,                  Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,                  Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,                  Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,                  Weinstein,G., and Gibbs,R.             </p> <p>                 Direct Submission                  Unpublished                  2 (bases 1 to 165679)                  Worley,K.C.                  Direct Submission                  Submitted (19-JUL-2002) Human Genome Sequencing Center, Department                  of Molecular and Human Genetics, Baylor College of Medicine, One                  Baylor Plaza, Houston, TX 77030, USA             </p> <p>                 Genome Center                  Center: Baylor College of Medicine                  Center code: BCM                  Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>                  Contact: hgsc-help@bcm.tmc.edu                  Project Information                  Center project name: GYGO                  Center clone name: CH230-402010             </p> <p>                 Summary Statistics                  Sequencing vector: Plasmid:                  Chemistry: Dye-terminator Big Dye: 100% of reads                  Assembly program: Phrap; version 0.990329                  Consensus quality: 125227 bases at least Q40                  Consensus quality: 130633 bases at least Q30                  Consensus quality: 135139 bases at least Q20             </p> <p>                 NOTE: Estimated insert size may differ from sequence length                  (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_dirft_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_dirft_data.html</a>)                  NOTE: This is a 'working draft' sequence. It currently                  consists of 41 contigs. The true order of the pieces                  is not known and their order in this sequence record is                  arbitrary. Gaps between the contigs are represented as                  runs of N, but the exact sizes of the gaps are unknown.                  This record will be updated with the finished sequence                  as soon as it is available and the accession number will                  be preserved.             </p> <table border="1"> <tr> <td>1</td> <td>1038</td> <td>contig of 1038 bp in length</td> </tr> <tr> <td>1039</td> <td>1138</td> <td>gap of unknown length</td> </tr> <tr> <td>1139</td> <td>2164</td> <td>contig of 1026 bp in length</td> </tr> <tr> <td>2165</td> <td>2264</td> <td>gap of unknown length</td> </tr> <tr> <td>2265</td> <td>3310</td> <td>contig of 1046 bp in length</td> </tr> <tr> <td>3311</td> <td>3410</td> <td>gap of unknown length</td> </tr> <tr> <td>3411</td> <td>5141</td> <td>contig of 1731 bp in length</td> </tr> <tr> <td>5142</td> <td>5241</td> <td>gap of unknown length</td> </tr> <tr> <td>5242</td> <td>6596</td> <td>contig of 1357 bp in length</td> </tr> <tr> <td>6599</td> <td>6698</td> <td>gap of unknown length</td> </tr> <tr> <td>6699</td> <td>7797</td> <td>contig of 1099 bp in length</td> </tr> <tr> <td>7798</td> <td>7897</td> <td>gap of unknown length</td> </tr> <tr> <td>7898</td> <td>9485</td> <td>contig of 1588 bp in length</td> </tr> <tr> <td>9486</td> <td>9585</td> <td>gap of unknown length</td> </tr> <tr> <td>9586</td> <td>11209</td> <td>contig of 1624 bp in length</td> </tr> <tr> <td>11210</td> <td>11309</td> <td>gap of unknown length</td> </tr> <tr> <td>11310</td> <td>13038</td> <td>contig of 1729 bp in length</td> </tr> <tr> <td>13039</td> <td>13138</td> <td>gap of unknown length</td> </tr> <tr> <td>13139</td> <td>14775</td> <td>contig of 1637 bp in length</td> </tr> <tr> <td>14776</td> <td>14875</td> <td>gap of unknown length</td> </tr> <tr> <td>14876</td> <td>16385</td> <td>contig of 1510 bp in length</td> </tr> <tr> <td>16386</td> <td>16485</td> <td>gap of unknown length</td> </tr> </table>	1	1038	contig of 1038 bp in length	1039	1138	gap of unknown length	1139	2164	contig of 1026 bp in length	2165	2264	gap of unknown length	2265	3310	contig of 1046 bp in length	3311	3410	gap of unknown length	3411	5141	contig of 1731 bp in length	5142	5241	gap of unknown length	5242	6596	contig of 1357 bp in length	6599	6698	gap of unknown length	6699	7797	contig of 1099 bp in length	7798	7897	gap of unknown length	7898	9485	contig of 1588 bp in length	9486	9585	gap of unknown length	9586	11209	contig of 1624 bp in length	11210	11309	gap of unknown length	11310	13038	contig of 1729 bp in length	13039	13138	gap of unknown length	13139	14775	contig of 1637 bp in length	14776	14875	gap of unknown length	14876	16385	contig of 1510 bp in length	16386	16485	gap of unknown length	<p>                 Maheshwari,M., Mapua,P., Martin,R., Mattindale,A., Martinez,E.,                  Massey,E., Mawlini,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,                  Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,                  Nguyen,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,                  Nguyen,N., Nickerson,E., Nwokweto,S., Ogun,M., Okunolu,G.,                  Oragunye,N., Oyler,R., Pace,A., Payton,B., Peery,J., Perez,L.,                  Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y.,                  Rives,M., Rojas,A., Kojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,                  Scherer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,I.,                  Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,                  Sutton,A., Swakek,A., Tabori,P., Tamerisa,A., Tameris,K., Tang,H.,                  Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,                  Usmani,R., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q.,                  Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,                  Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,                  Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,                  Weinstein,G., and Gibbs,R.             </p> <p>                 Direct Submission                  Unpublished                  2 (bases 1 to 165679)                  Worley,K.C.                  Direct Submission                  Submitted (19-JUL-2002) Human Genome Sequencing Center, Department                  of Molecular and Human Genetics, Baylor College of Medicine, One                  Baylor Plaza, Houston, TX 77030, USA             </p> <p>                 Genome Center                  Center: Baylor College of Medicine                  Center code: BCM                  Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>                  Contact: hgsc-help@bcm.tmc.edu                  Project Information                  Center project name: GYGO                  Center clone name: CH230-402010             </p> <p>                 Summary Statistics                  Sequencing vector: Plasmid:                  Chemistry: Dye-terminator Big Dye: 100% of reads                  Assembly program: Phrap; version 0.990329                  Consensus quality: 125227 bases at least Q40                  Consensus quality: 130633 bases at least Q30                  Consensus quality: 135139 bases at least Q20             </p> <p>                 NOTE: Estimated insert size may differ from sequence length                  (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_dirft_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_dirft_data.html</a>)                  NOTE: This is a 'working draft' sequence. It currently                  consists of 41 contigs. The true order of the pieces                  is not known and their order in this sequence record is                  arbitrary. Gaps between the contigs are represented as                  runs of N, but the exact sizes of the gaps are unknown.                  This record will be updated with the finished sequence                  as soon as it is available and the accession number will                  be preserved.             </p> <table border="1"> <tr> <td>1</td> <td>1038</td> <td>contig of 1038 bp in length</td> </tr> <tr> <td>1039</td> <td>1138</td> <td>gap of unknown length</td> </tr> <tr> <td>1139</td> <td>2164</td> <td>contig of 1026 bp in length</td> </tr> <tr> <td>2165</td> <td>2264</td> <td>gap of unknown length</td> </tr> <tr> <td>2265</td> <td>3310</td> <td>contig of 1046 bp in length</td> </tr> <tr> <td>3311</td> <td>3410</td> <td>gap of unknown length</td> </tr> <tr> <td>3411</td> <td>5141</td> <td>contig of 1731 bp in length</td> </tr> <tr> <td>5142</td> <td>5241</td> <td>gap of unknown length</td> </tr> <tr> <td>5242</td> <td>6596</td> <td>contig of 1357 bp in length</td> </tr> <tr> <td>6599</td> <td>6698</td> <td>gap of unknown length</td> </tr> <tr> <td>6699</td> <td>7797</td> <td>contig of 1099 bp in length</td> </tr> <tr> <td>7798</td> <td>7897</td> <td>gap of unknown length</td> </tr> <tr> <td>7898</td> <td>9485</td> <td>contig of 1588 bp in length</td> </tr> <tr> <td>9486</td> <td>9585</td> <td>gap of unknown length</td> </tr> <tr> <td>9586</td> <td>11209</td> <td>contig of 1624 bp in length</td> </tr> <tr> <td>11210</td> <td>11309</td> <td>gap of unknown length</td> </tr> <tr> <td>11310</td> <td>13038</td> <td>contig of 1729 bp in length</td> </tr> <tr> <td>13039</td> <td>13138</td> <td>gap of unknown length</td> </tr> <tr> <td>13139</td> <td>14775</td> <td>contig of 1637 bp in length</td> </tr> <tr> <td>14776</td> <td>14875</td> <td>gap of unknown length</td> </tr> <tr> <td>14876</td> <td>16385</td> <td>contig of 1510 bp in length</td> </tr> <tr> <td>16386</td> <td>16485</td> <td>gap of unknown length</td> </tr> </table>	1	1038	contig of 1038 bp in length	1039	1138	gap of unknown length	1139	2164	contig of 1026 bp in length	2165	2264	gap of unknown length	2265	3310	contig of 1046 bp in length	3311	3410	gap of unknown length	3411	5141	contig of 1731 bp in length	5142	5241	gap of unknown length	5242	6596	contig of 1357 bp in length	6599	6698	gap of unknown length	6699	7797	contig of 1099 bp in length	7798	7897	gap of unknown length	7898	9485	contig of 1588 bp in length	9486	9585	gap of unknown length	9586	11209	contig of 1624 bp in length	11210	11309	gap of unknown length	11310	13038	contig of 1729 bp in length	13039	13138	gap of unknown length	13139	14775	contig of 1637 bp in length	14776	14875	gap of unknown length	14876	16385	contig of 1510 bp in length	16386	16485	gap of unknown length
1	1038	contig of 1038 bp in length																																																																																																																																			
1039	1138	gap of unknown length																																																																																																																																			
1139	2164	contig of 1026 bp in length																																																																																																																																			
2165	2264	gap of unknown length																																																																																																																																			
2265	3310	contig of 1046 bp in length																																																																																																																																			
3311	3410	gap of unknown length																																																																																																																																			
3411	5141	contig of 1731 bp in length																																																																																																																																			
5142	5241	gap of unknown length																																																																																																																																			
5242	6596	contig of 1357 bp in length																																																																																																																																			
6599	6698	gap of unknown length																																																																																																																																			
6699	7797	contig of 1099 bp in length																																																																																																																																			
7798	7897	gap of unknown length																																																																																																																																			
7898	9485	contig of 1588 bp in length																																																																																																																																			
9486	9585	gap of unknown length																																																																																																																																			
9586	11209	contig of 1624 bp in length																																																																																																																																			
11210	11309	gap of unknown length																																																																																																																																			
11310	13038	contig of 1729 bp in length																																																																																																																																			
13039	13138	gap of unknown length																																																																																																																																			
13139	14775	contig of 1637 bp in length																																																																																																																																			
14776	14875	gap of unknown length																																																																																																																																			
14876	16385	contig of 1510 bp in length																																																																																																																																			
16386	16485	gap of unknown length																																																																																																																																			
1	1038	contig of 1038 bp in length																																																																																																																																			
1039	1138	gap of unknown length																																																																																																																																			
1139	2164	contig of 1026 bp in length																																																																																																																																			
2165	2264	gap of unknown length																																																																																																																																			
2265	3310	contig of 1046 bp in length																																																																																																																																			
3311	3410	gap of unknown length																																																																																																																																			
3411	5141	contig of 1731 bp in length																																																																																																																																			
5142	5241	gap of unknown length																																																																																																																																			
5242	6596	contig of 1357 bp in length																																																																																																																																			
6599	6698	gap of unknown length																																																																																																																																			
6699	7797	contig of 1099 bp in length																																																																																																																																			
7798	7897	gap of unknown length																																																																																																																																			
7898	9485	contig of 1588 bp in length																																																																																																																																			
9486	9585	gap of unknown length																																																																																																																																			
9586	11209	contig of 1624 bp in length																																																																																																																																			
11210	11309	gap of unknown length																																																																																																																																			
11310	13038	contig of 1729 bp in length																																																																																																																																			
13039	13138	gap of unknown length																																																																																																																																			
13139	14775	contig of 1637 bp in length																																																																																																																																			
14776	14875	gap of unknown length																																																																																																																																			
14876	16385	contig of 1510 bp in length																																																																																																																																			
16386	16485	gap of unknown length																																																																																																																																			

\* 16486 18835: contig of 2350 bp in length  
\* 18836 18935: gap of unknown length  
\* 18936 21757: contig of 2822 bp in length  
\* 21758 21857: gap of unknown length  
\* 21858 24748: contig of 2891 bp in length  
\* 24749 24848: gap of unknown length  
\* 24849 27365: contig of 2717 bp in length  
\* 27366 30397: contig of 2732 bp in length  
\* 30398 30498: gap of unknown length  
\* 30498 34133: contig of 3636 bp in length  
\* 34134 34233: gap of unknown length  
\* 34234 36775: contig of 2342 bp in length  
\* 36776 36875: gap of unknown length  
\* 36876 40849: contig of 3974 bp in length  
\* 40850 40949: gap of unknown length  
\* 40950 44518: contig of 3569 bp in length  
\* 44519 44618: gap of unknown length  
\* 44619 48721: contig of 4103 bp in length  
\* 48722 48821: gap of unknown length  
\* 48822 52408: contig of 3587 bp in length  
\* 52409 52508: gap of unknown length  
\* 52509 55637: contig of 3129 bp in length  
\* 55638 55737: gap of unknown length  
\* 55738 59567: contig of 3830 bp in length  
\* 59568 59667: gap of unknown length  
\* 59668 62468: contig of 2801 bp in length  
\* 62469 62568: gap of unknown length  
\* 62569 66315: contig of 3747 bp in length  
\* 66316 66415: gap of unknown length  
\* 66416 70295: contig of 3860 bp in length  
\* 70296 70395: gap of unknown length  
\* 70396 75318: contig of 4923 bp in length  
\* 75319 75418: gap of unknown length  
\* 75419 81215: contig of 5797 bp in length  
\* 81216 81315: gap of unknown length  
\* 81316 86210: contig of 4895 bp in length  
\* 86211 86310: gap of unknown length  
\* 86311 92154: contig of 5844 bp in length  
\* 92155 92254: gap of unknown length  
\* 92255 97004: contig of 4750 bp in length  
\* 97005 97104: gap of unknown length  
\* 97105 103480: contig of 6376 bp in length  
\* 103481 103580: gap of unknown length  
\* 103581 108040: contig of 4460 bp in length  
\* 108041 108140: gap of unknown length  
\* 108141 113761: contig of 5621 bp in length  
\* 113762 113861: gap of unknown length  
\* 113862 118463: contig of 4602 bp in length  
\* 118464 118563: gap of unknown length  
\* 118564 127494: contig of 8931 bp in length  
\* 127495 127594: gap of unknown length  
\* 127595 134922: contig of 7328 bp in length  
\* 134923 135022: gap of unknown length  
\* 135023 144536: contig of 9514 bp in length  
\* 144537 144636: gap of unknown length  
\* 144637 155238: contig of 10602 bp in length  
\* 155239 155338: gap of unknown length  
\* 155339 165679: contig of 10341 bp in length.

## FEATURES

source

1.165679

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/clone="CH230-402010"

BASE COUNT 50332 a 29730 c 29146 g 47927 t 8544 others

## ORIGIN

Query Match

Best Local Similarity 100.0%; Score 21; DB 2; Length 165679;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1576 ATTGAGCTCTGCTTCTGAT 1596  
|||||

Db 53196 ATTCAGCTCTGCTTCTGAT 53216

## RESULT 22

AP001554/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

BASE COUNT 44526 a 43542 c 41318 g 42361 t

## ORIGIN

Query Match

Best Local Similarity 100.0%; Score 21; DB 9; Length 171747;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 446 CCTACTCCAGATTGCTAGG 466  
|||||

Db 94200 CCTACTCCAGATTGCTAGG 94180

## RESULT 23

AC020766

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (21-AUG-2002) Production Sequencing Facility, DOE Joint



```

* 79941 80040: gap of 100 bp
* 80041 91399: contig of 11359 bp in length
* 91400 91499: gap of 100 bp
* 91500 102181: contig of 10682 bp in length
* 102182 102281: gap of 100 bp
* 102282 114041: contig of 11760 bp in length
* 114042 114141: gap of 100 bp
* 114142 125232: contig of 11091 bp in length
* 125233 125332: gap of 100 bp
* 125333 132700: contig of 7368 bp in length
* 132701 132800: gap of 100 bp
* 132801 140074: contig of 7274 bp in length
* 140075 140174: gap of 100 bp
* 140175 145154: contig of 4980 bp in length
* 145155 145254: gap of 100 bp
* 145255 150197: contig of 4943 bp in length
* 150198 150297: gap of 100 bp
* 150298 154284: contig of 3987 bp in length
* 154285 154384: gap of 100 bp
* 154385 158901: contig of 4517 bp in length
* 158902 159001: gap of 100 bp
* 159002 162569: contig of 3568 bp in length
* 162570 162669: gap of 100 bp
* 162670 166056: contig of 3387 bp in length
* 166057 166156: gap of 100 bp
* 166157 169174: contig of 3018 bp in length
* 169175 169274: gap of 100 bp
* 169275 171034: contig of 1760 bp in length
* 171035 171134: gap of 100 bp
* 171135 173192: contig of 2058 bp in length
* 173193 173292: gap of 100 bp
* 173293 175034: contig of 1742 bp in length
* 175035 175134: gap of 100 bp
* 175135 176781: contig of 1647 bp in length
* 176782 176881: gap of 100 bp
* 176882 178330: contig of 1449 bp in length
* 178331 178430: gap of 100 bp
* 178431 180012: contig of 1582 bp in length
* 180013 180112: gap of 100 bp
* 180113 181489: contig of 1377 bp in length
* 181490 181589: gap of 100 bp
* 181590 183286: contig of 1699 bp in length
* 183289 183388: gap of 100 bp
* 183389 184908: contig of 1520 bp in length
* 184909 185008: gap of 100 bp
* 185009 186429: contig of 1421 bp in length
* 186430 186529: gap of 100 bp
* 186530 187961: contig of 1432 bp in length
* 187962 188061: gap of 100 bp
* 188062 189626: contig of 1565 bp in length
* 189627 189726: gap of 100 bp
* 189727 191063: contig of 1337 bp in length
* 191064 191163: gap of 100 bp
* 191164 192407: contig of 1244 bp in length
* 192408 192507: gap of 100 bp
* 192508 193854: contig of 1347 bp in length
* 193855 193954: gap of 100 bp
* 193955 195326: contig of 1372 bp in length
* 195327 195426: gap of 100 bp
* 195427 196485: contig of 1059 bp in length
* 196486 196585: gap of 100 bp
* 196586 197657: contig of 1072 bp in length
* 197658 197757: gap of 100 bp
* 197758 199077: contig of 1320 bp in length
* 199078 199177: gap of 100 bp
* 199178 200305: contig of 1128 bp in length
* 200306 200405: gap of 100 bp
* 200406 201591: contig of 1186 bp in length
* 201592 201691: gap of 100 bp
* 201692 202867: contig of 1176 bp in length
* 202868 202967: gap of 100 bp
* 202968 204235: contig of 1268 bp in length
* 204236 204335: gap of 100 bp

```

```

* 204336 205348: contig of 1013 bp in length
* 205349 205448: gap of 100 bp
* 205449 206463: contig of 1015 bp in length.
FEATURES
  source
    1..206463
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="11"
      /clone="RP11-853A16"
      1..34031
        /note="assembly_fragment"
      34132..48505
        /note="assembly_fragment"
      48606..62211
        /note="assembly_fragment"
      62312..77940
        /note="assembly_fragment"
misc-feature
  misc-feature
    1..34031
      /note="assembly_fragment"
    34132..48505
      /note="assembly_fragment"
    48606..62211
      /note="assembly_fragment"
    62312..77940
      /note="assembly_fragment"
Query Match
  Best Local Similarity 100.0%; Score 21; DB 2; Length 206463;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 446 CCGACCTCCAGGATTGCTAGG 466
Db 21073 CCGACCTCCAGGATTGCTAGG 21093
RESULT 25
AX098422
LOCUS
  AX098422 1770 bp mRNA linear PAT 02-APR-2001
DEFINITION
  Sequence 1 from Patent WO0120009.
ACCESSION
  AX098422
VERSION
  AX098422.1 GI:13537714
KEYWORDS
  SOURCE
    thale cress.
  ORGANISM
    Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
REFERENCE
  1 (bases 1 to 1770)
  Reindl, A., Geigenberger, P. L., Neuhaus, H. E., Graeve-Kampfenkel, K.,
  Moehlmann, T. and Tjaden, J.
  Plants having altered amino acid contents and method for the
  production thereof
  Patent: WO 0120009-A 1 22-MAR-2001;
JOURNAL
  BASF AKTIENGESELLSCHAFT (DE)
FEATURES
  source
    1..1770
      /organism="Arabidopsis thaliana"
      /db_xref="taxon:3702"
BASE COUNT
  441 a 383 c 431 g 515 t
ORIGIN
Query Match
  Best Local Similarity 100.0%; Score 20; DB 6; Length 1770;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1308 AAGAGAAATGGCTATATGCC 1327
Db 1489 AAGAGAAATGGCTATATGCC 1508
RESULT 26
AF428316
LOCUS
  AF428316 2110 bp mRNA linear PLN 18-OCT-2001
DEFINITION
  Arabidopsis thaliana At1g80300/F516_5 mRNA, complete cds.
ACCESSION
  AF428316
VERSION
  AF428316.1 GI:16226733
KEYWORDS
  F11.CDNA.
  Arabidopsis thaliana.
  Arabidopsis thaliana.
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

REFERENCE  
AUTHORS

Spermatophyta: Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 2110)

Chen, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamliya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Arabidopsis

Unpublished

2 (bases 1 to 2110)

Chen, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamliya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE  
JOURNAL

Direct Submission  
Submitted (05-OCT-2001) Salk Institute Genomic Analysis Laboratory (Signal), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

## COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamliya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Chen, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Shin, P., Banh, J., Bowser, L., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.

Chen, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES  
SOURCE

Location/Qualifiers

1. 2110

/organism="Arabidopsis thaliana"

/db\_xref="taxon:3702"

/chromosome="1"

/clone="RAFL09-51-CL18(R14701)"

/note="ecotype: Columbia"

1. 55

56. 1930

/note="adenine nucleotide translocase"

/codon\_start=1

/product="Atlg80300/F516-5"

/protein\_id="AAL16246.1"

/db\_xref="GI:16226734"

/translation="MEAVIQIRGLSLPTKPIGVRSOLQPSHGILKORLFAAPRNILG LSLFNGHKRFQEPETHGISHSKERSTFICKAEAAAGDVAEGEGSAVAVS PKIIGVAVATLKIIIPLGMPICILNTYTRDKDVLVAVAKASITPEELTWN LPMAGFMILYTKSLNYLSKALPFYIVIPRTYFGAFGPNYPLSNTIHEALADKL LTTIGPRMPGFIATLRIMSCLFYVMAFLMGSVAVSLFNGFAQITTVDAKFTPL FGLGAVNALPFSGRVTKYFSNLRNKLDPGVDGVAVSLKAMMSIVGMLACILYMW NRYVLPDTRSKNKKPKMKPMESLKLVSPPYIRDLATLVAVGISINLVEVWKS LKQFSPSPNSAFMGDFSTCTGVATFTMMLDSQYFNKVGVAAKTTPVLILTVY AEFSLIFGFPFADLVAKLGMTPLLAAYVVALONIFSKSAKSYSLPDCCKMAYIPD EDPYVKRAIDVYCNPLKSGALIOQFMILSRGSLANSPPYGMILYIVTMILAA AKSLGQPNLSLRSEELKEMERASSVKIPVDSDESGNSLGESPSSPKSPKAPTNL

5'UTR  
CDS

/translation="MEAVIQIRGLSLPTKPIGVRSOLQPSHGILKORLFAAPRNILG LSLFNGHKRFQEPETHGISHSKERSTFICKAEAAAGDVAEGEGSAVAVS PKIIGVAVATLKIIIPLGMPICILNTYTRDKDVLVAVAKASITPEELTWN LPMAGFMILYTKSLNYLSKALPFYIVIPRTYFGAFGPNYPLSNTIHEALADKL LTTIGPRMPGFIATLRIMSCLFYVMAFLMGSVAVSLFNGFAQITTVDAKFTPL FGLGAVNALPFSGRVTKYFSNLRNKLDPGVDGVAVSLKAMMSIVGMLACILYMW NRYVLPDTRSKNKKPKMKPMESLKLVSPPYIRDLATLVAVGISINLVEVWKS LKQFSPSPNSAFMGDFSTCTGVATFTMMLDSQYFNKVGVAAKTTPVLILTVY AEFSLIFGFPFADLVAKLGMTPLLAAYVVALONIFSKSAKSYSLPDCCKMAYIPD EDPYVKRAIDVYCNPLKSGALIOQFMILSRGSLANSPPYGMILYIVTMILAA AKSLGQPNLSLRSEELKEMERASSVKIPVDSDESGNSLGESPSSPKSPKAPTNL

/translation="MEAVIQIRGLSLPTKPIGVRSOLQPSHGILKORLFAAPRNILG LSLFNGHKRFQEPETHGISHSKERSTFICKAEAAAGDVAEGEGSAVAVS PKIIGVAVATLKIIIPLGMPICILNTYTRDKDVLVAVAKASITPEELTWN LPMAGFMILYTKSLNYLSKALPFYIVIPRTYFGAFGPNYPLSNTIHEALADKL LTTIGPRMPGFIATLRIMSCLFYVMAFLMGSVAVSLFNGFAQITTVDAKFTPL FGLGAVNALPFSGRVTKYFSNLRNKLDPGVDGVAVSLKAMMSIVGMLACILYMW NRYVLPDTRSKNKKPKMKPMESLKLVSPPYIRDLATLVAVGISINLVEVWKS LKQFSPSPNSAFMGDFSTCTGVATFTMMLDSQYFNKVGVAAKTTPVLILTVY AEFSLIFGFPFADLVAKLGMTPLLAAYVVALONIFSKSAKSYSLPDCCKMAYIPD EDPYVKRAIDVYCNPLKSGALIOQFMILSRGSLANSPPYGMILYIVTMILAA AKSLGQPNLSLRSEELKEMERASSVKIPVDSDESGNSLGESPSSPKSPKAPTNL

1931. 2110

454 c

502 g

629 t

3'UTR  
BASE COUNT

525 a

454 c

502 g

629 t

## ORIGIN

Query Match 1.2%; Score 20; DB 8; Length 2110; Best Local Similarity 100.0%; Pred. No. 38; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1308 AAGCAATGCGCTATATCC 1327

Db 1553 AAGCAATGCGCTATATCC 1572

RESULT 27  
AY045903  
LOCUS

DEFINITION

Arabidopsis thaliana putative adenine nucleotide translocase (Atlg80300) mRNA, complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
AUTHORS

1 (bases 1 to 2151)

Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamliya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shin, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Arabidopsis Full Length cDNA Clones

Unpublished

2 (bases 1 to 2151)

Yamada, K., Liu, S.X., Pham, P.K., Banh, J., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamliya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shin, P., Southwick, A., Tracy, S.E., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Direct Submission

Submitted (10-JUL-2001)

Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamliya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

COMMENT

LOCATION/QUALIFIERS

1. 2151

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X., Pham, P.K., Banh, J., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shin, P., Southwick, A., Tracy, S.E., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

FEATURES  
SOURCE

Location/Qualifiers

1. 2151

/organism="Arabidopsis thaliana"

/db\_xref="taxon:3702"

/chromosome="1"

/clone="RAFL07-08-J19 (R10784)"

/note="This clone is in a modified pBluescript vector (PIC-1) as a BamHI/XhoI insert.

ecotype: Columbia"

```

gene      1. .2151
          /gene="Atlg80300"
5'UTR     1..27
          /gene="Atlg80300"
CDS       28..1902
          /gene="Atlg80300"
          /codon_start=1
          /evidence=experimental
          /product="putative adenine nucleotide translocase"
          /protein_id="AAK76577.1"
          /db_xref="GI:15028093"
          /translation="MEAVIOTRGLSLPTKPIGVRSQLOPSHGLKQRLFAAKPRNLHG
          LSLFNGHKKFOTPEPTLHGISISHKRSTERICAEAAAGDGAAPGEGSAAVAAS
          PKRFEVATLTKIIPGLMFCILFNTYIILDQDVLVYAKSSAEIIPPLKTVNL
          LPMAGFMULTKLSNVLSKALFTVIVPITFGAFGVNPLSNITVDEAKRYP
          LTTGPRMGPPIAIRISFCLFYVMAELGWSVVSVLFMGFANOITTVDEAKRYP
          FLGGNVALIFSGRTVKYSNLRKMLGVDGMAVSLKAMSIYVGMICLILYMW
          NRYVPLPTRSKNKKPKMGKTESLKLPLVSSPYRIDATLVAAVGISINLVEVWKS
          LKAOFPSPNEXSAFMGDSCTGVAFTFMMLLSOVFNKYGVAAKITPVVLTGV
          AFPSLILFGPPAPLAKLGMTPILAAVYALONIFSKAKSYSLFDECKEMAYIPLD
          EDTYVKGRALIDVVCNPLGKSGALIQPFMLISFSLANSIPYLGMILLVITAMLA
          AKSLGQPNLSRSEEELEKEMERASSVKIPVVSQDSNGSLGESPPSSPEKSAFTNL"

misc_difference 869
          /gene="Atlg80300"
          /note="compared to genomic sequence resulting in an amino
          acid sequence difference"
          /replace="C"
          3'UTR     1903..2151
          /gene="Atlg80300"
          /gene="Atlg80300"
          /note="compared to genomic sequence"
          /replace="C"

BASE COUNT  553 a      451 c      510 g      637 t
ORIGIN
Query Match      1..2%: Score 20; DB 8; Length 2151;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1308 AAAGAAATGGCCTATATATCC 1327
      |||||||||||||||||||
Db 1525 AAAGAAATGGCCTATATATCC 1544

RESULT 28
ATATMTR      2181 bp      mRNA      linear      PLN 26-NOV-1999
DEFINITION   A.thaliana mRNA for adenine nucleotide translocase.
ACCESSION   Z49227.2
VERSION      GI:6469339
KEYWORDS     adenine nucleotide translocase.
SOURCE       Arabidopsis thaliana.
ORGANISM     Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 2181)
Kampfenkel,K., Mohlmann,T., Batz,O., Van Montagu,M., Inze,D. and
Neuhaus,H.E.
REFERENCE    1
AUTHORS      Kampfenkel,K., Mohlmann,T., Batz,O., Van Montagu,M., Inze,D. and
Neuhaus,H.E.
TITLE        Molecular characterization of an Arabidopsis thaliana cDNA encoding
a novel putative adenylate translocator of higher plants
JOURNAL      FEBS Lett. 374 (3), 351-355 (1995)
MEDLINE      96069943
PUBMED       7589569
REMARK       (sites)
REFERENCE    2 (bases 1 to 2181)
AUTHORS      Kampfenkel,K.K.
TITLE        Direct Submission
JOURNAL      Submitted (05-MAY-1995) Kampfenkel K.K., Universiteit Gent,
Laboratorium voor Genetika, K.L. Ledeganckstraat 35, Gent, Belgium,
B-9000 Gent

```

```

REMARK      revised by [3]
REFERENCE    3 (bases 1 to 2181)
AUTHORS      Kampfenkel,K.K.
TITLE        Direct Submission
JOURNAL      Submitted (26-NOV-1999) Kampfenkel K.K., Universiteit Gent,
Laboratorium voor Genetika, K.L. Ledeganckstraat 35, Gent, Belgium,
B-9000 Gent
COMMENT      On Nov 27, 1999 this sequence version replaced gi:1051108.
FEATURES
source
1..2181
/organism="Arabidopsis thaliana"
/strain="var. Columbia"
/db_xref="taxon:3702"
/clone="C23"
/tissue="type="whole seedling"
/clone_lib="Lambdages"
/dev_stage="seedling"
110..1981
/codon_start=1
/product="adenine nucleotide translocase"
/protein_id="CA89201.2"
/db_xref="GI:6469340"
/db_xref="SPTREMBL:Q39002"
/translation="MEAVIOTRGLSLPTKPIGVRSQLOPSHGLKQRLFAAKPRNLHG
CLYPLTGTRNRPUSQPCMGFRFPTRKRAPSSVARRRGRCWRSCLRSDSAVAASR
KIFGEVATLTKIIPGLMFCILFNTYIILDQDVLVYAKSSAEIIPPLKTVNL
LPMAGFMULTKLSNVLSKALFTVIVPITFGAFGVNPLSNITVDEAKRYP
LTTGPRMGPPIAIRISFCLFYVMAELGWSVVSVLFMGFANOITTVDEAKRYP
GLGANVALIFSGRTVKYSNLRKMLGVDGMAVSLKAMSIYVGMICLILYMW
RYVPLPTRSKNKKPKMGKTESLKLPLVSSPYRIDATLVAAVGISINLVEVWKS
LKAOFPSPNEXSAFMGDSCTGVAFTFMMLLSOVFNKYGVAAKITPVVLTGV
AFPSLILFGPPAPLAKLGMTPILAAVYALONIFSKAKSYSLFDECKEMAYIPLD
EDTYVKGRALIDVVCNPLGKSGALIQPFMLISFSLANSIPYLGMILLVITAMLA
AKSLGQPNLSRSEEELEKEMERASSVKIPVVSQDSNGSLGESPPSSPEKSAFTNL"

BASE COUNT  549 a      476 c      508 g      648 t
ORIGIN
Query Match      1..2%: Score 20; DB 8; Length 2181;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1308 AAAGAAATGGCCTATATATCC 1327
      |||||||||||||||||||
Db 1604 AAAGAAATGGCCTATATATCC 1623

RESULT 29
AF222985
LOCUS       AF222985      2762 bp      DNA      linear      PRI 02-JUN-2000
DEFINITION   Homo sapiens derived chromosome 1 translocation breakpoint
sequence.
ACCESSION   AF222985
VERSION      AF222985.1
KEYWORDS     GI:8163838
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2762)
Miller,J.K., Wilson-Annan,J.C., Anderson,S., Christie,S.,
Taylor,M.S., Sempile,C.A., Devon,R.S., Clair,D.M., Muir,W.J.,
Blackwood,D.H. and Porteous,D.J.
Disruption of two novel genes by a translocation co-segregating
with schizophrenia
Hum. Mol. Genet. 9 (9), 1415-1423 (2000)
JOURNAL      20275630
MEDLINE      10814723
PUBMED       10814723
REFERENCE    2 (bases 1 to 2762)
AUTHORS      Miller,K., Anderson,S. and Christie,S.
TITLE        Direct Submission
JOURNAL      Submitted (10-JAN-2000) Medical Genetics Section, The University of
Edinburgh, Crewe Road, Edinburgh EH4 2XU, Scotland
FEATURES
Location/Qualifiers

```

CDS	<5357..>5545 /gene="DISC1" /codon_start=3 /product="DISC1 protein" /protein_id="AA073874.1" /db_xref="GI:8163836" /translation="NH2MTAKDLTEEFIRSLTSEREGGLSKLVISNRNKKLGSV KEIDNRIRREVERHETAY" /cds_start=5357.. /cds_end=5545 /number=9
exon	
BASE COUNT	1968 a 1354 c 1430 g 2526 t 13 others
ORIGIN	
Query Match	1.2%; Score 20; DB 9; Length 7291;
Best Local Similarity	100.0%; Pred. No. 37;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	913 TTTTATTACTCTCTGGTT 932 
Db	339 TTTTATTACTCTCTGGTT 358
RESULT 31	
LOCUS	AF222981 15002 bp DNA linear PRI 02-JUN-2000
DEFINITION	Homo sapiens 1 DISC2 gene, complete sequence.
ACCESSION	AF222981
VERSION	AF222981.1 GI:8163870
KEYWORDS	
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 15002) Miller,J.K., Wilson-Annan,J.C., Anderson,S., Christie,S., Taylor,M.S., Sempke,C.A., Devon,R.S., Clair,D.M., Muir,W.J., Blackwood,D.H. and Porteous,D.J. Disruption of two novel genes by a translocation co-segregating with schizophrenia Hum. Mol. Genet. 9 (9), 1415-1423 (2000)
TITLE	
JOURNAL	
MEDLINE	20275630
PUBMED	10814723
REFERENCE	2 (bases 1 to 15002) Miller,K., Anderson,S. and Christie,S. Direct Submission Submitted (11-JUN-2000) Medical Genetics Section, The University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, Scotland
JOURNAL	
TITLE	
FEATURES	Location/Qualifiers 1..15002 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /map="1q42.1" 1..15002 /gene="DISC2" /note="considered to be a non-coding; 5' end has not yet been identified"
BASE COUNT	4815 a 3132 c 2838 g 4209 t 8 others
ORIGIN	
Query Match	1.2%; Score 20; DB 9; Length 15002;
Best Local Similarity	100.0%; Pred. No. 37;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	913 TTTTATTACTCTCTGGTT 932 
Db	11647 TTTTATTACTCTCTGGTT 11628
RESULT 32	
LOCUS	AF222987 33376 bp DNA linear PRI 02-JUN-2000
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
JOURNAL	
TITLE	
FEATURES	
BASE COUNT	
ORIGIN	

DEFINITION Homo sapiens DISC1 protein (DISC1) gene, partial cds and DISC2  
gene, partial sequence.  
ACCESSION AF222987  
VERSION AF222987.1 GI:8163844  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 33376)  
AUTHORS Miller,J.K., Wilson-Annan,J.C., Anderson,S., Christie,S.,  
Taylor,M.S., Semple,C.A., Devon,R.S., Clair,D.M., Mulr,W.J.,  
Blackwood,D.H. and Porteous,D.J.  
TITLE Disruption of two novel genes by a translocation co-segregating  
with schizophrenia  
Hum. Mol. Genet. 9 (9), 1415-1423 (2000)  
JOURNAL MEDLINE  
MEDLINE 20275630  
PUBMED 10814723  
REFERENCE 2 (bases 1 to 33376)  
AUTHORS Anderson,S., Miller,K. and Christie,S.  
TITLE Direct Submission  
JOURNAL Submitted (11-JAN-2000) Medical Genetics Section, The University of  
Edinburgh, Crewe Road, Edinburgh EH4 2XU, Scotland  
FEATURES  
Source Location/Qualifiers  
1..33376  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q42.1"  
1..33376  
misc\_feature  
1..8206  
/note="translocation breakpoint region"  
misc\_feature  
1..8206  
/note="contig1"  
CDS <4771..>26773  
/gene="DISC1"  
join(<4771..4825,8413..8515,26585..>26773)  
/gene="DISC1"  
/codon\_start=2  
/product="DISC1 protein"  
/protein\_id="AAF73877.1"  
/db\_xref="GI:8163845"  
/translation="LQBRKISNLSKEITTKVMSRPFCSIRKKYNDIETQPAL  
EAKHAIISGHFWTKADTEIRISLTSEKGLSKLLVLSRNVKKGSKVEDYN  
RLRREVEHETAY"  
4771..4825  
/gene="DISC1"  
/number=7  
misc\_feature  
8206..33376  
/note="contig2"  
8413..8515  
/gene="DISC1"  
/number=8  
gene complement(18199..>33376)  
/gene="DISC2"  
exon complement(18199..>33376)  
/gene="DISC2"  
26585..26773  
/gene="DISC1"  
/number=9  
BASE COUNT 9071 a 6789 c 7155 g 10350 t 11 others  
ORIGIN  
Query Match 1.2%: Score 20; DB 9; Length 33376;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 913 TTTTATTACTCTCTGTGTT 932  
|||||  
Db 21554 TTTTATTACTCTCTGTGTT 21573  
RESULT 33  
AX151213

LOCUS AX151213 33780 bp DNA linear PAT 22-JUN-2001  
DEFINITION Sequence 4 from Patent WO0140301.  
ACCESSION AX151213  
VERSION AX151213.1 GI:14533379  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 33780)  
AUTHORS Porteous,D., Miller,K. and Blackwood,D.  
TITLE A gene, disrupted in schizophrenia  
JOURNAL Patent: WO 0140301-A 4 07-JUN-2001;  
Akzo Nobel N.V. (NL); MEDICAL RESEARCH COUNCIL (GB); UNIVERSITY  
OF EDINBURGH (GB)  
FEATURES  
Source Location/Qualifiers  
1..33780  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 9140 a 6838 c 7205 g 10408 t 189 others  
ORIGIN  
Query Match 1.2%: Score 20; DB 6; Length 33780;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 913 TTTTATTACTCTCTGTGTT 932  
|||||  
Db 21958 TTTTATTACTCTCTGTGTT 21977  
RESULT 34  
AC107059 48546 bp DNA linear PRI 20-APR-2002  
LOCUS AC107059  
DEFINITION Homo sapiens BAC clone RP11-668G10 from 4, complete sequence.  
ACCESSION AC107059  
VERSION AC107059.6 GI:19848496  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 48546)  
AUTHORS Sulston,J.E. and Waterston,R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 48546)  
AUTHORS Cedroni,M. and Cotton,M.  
TITLE The sequence of Homo sapiens BAC clone RP11-668G10  
JOURNAL Unpublished (2001)  
REFERENCE 3 (bases 1 to 48546)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (14-JAN-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 48546)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAR-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 5 (bases 1 to 48546)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 6 (bases 1 to 48546)  
AUTHORS Waterston,R.  
TITLE Direct Submission



JOURNAL  
Submitted (20-APR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Mar 30, 2002 this sequence version replaced gi:19551219.

## COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0668610

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frenken, E., Tatenno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-336N6, 2000 bp overlap; the clone sequenced to the right is RP11-469011, 2000 bp overlap.  
Actual start of this clone is at base position 123490 of RP11-336N6; actual end is at base position 104312 of RP11-469011.

Polymorphisms have been identified between AC107059, AC055120 and AC012504.

## FEATURES

SOURCE  
1. 48546  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-668610"  
/clone\_lib="RPCI-11"  
112. 127  
/note="match to EST AI654124 (NID:g4738103) ty62a05.x1"  
112. 125  
/note="similar to Homo sapiens EST AI741715 (NID:g5110003) wg22b05.x1"  
112. 123  
/note="match to EST AW449862 (NID:g6990638) "  
113. 124  
/note="match to EST AI962124 (NID:g5754837) wx77f01.x1"  
119. 438  
/note="similar to Homo sapiens EST H78148 (NID:g1056237) yu86f10.r1"  
122. 385  
/note="similar to Homo sapiens EST AA640399 (NID:g2565649) nt99e01.s1"  
misc\_feature  
126. 812

/note="similar to Homo sapiens EST BF669641 (NID:g11943536) "  
342. 352  
/note="match to EST AI962124 (NID:g5754837) wx77f01.x1"  
393. 749  
/note="match to EST AI969703 (NID:g5766521) wz70f08.x1"  
450. 848  
/note="match to EST AA863006 (NID:g2955485) og99d10.s1"  
546. 952  
/note="similar to EST BM541231 (NID:g18824674) "  
598. 1112  
/note="similar to Homo sapiens EST AI741715 (NID:g5110003) wg22b05.x1"  
600. 900  
/note="match to EST AA401689 (NID:g2057173) zt66a07.s1"  
658. 1140  
/note="similar to Homo sapiens EST AU525280 (NID:g12788773) "  
671. 1347  
/note="similar to Homo sapiens EST B1490709 (NID:g15329937) "  
674. 1364  
/note="similar to Homo sapiens EST AU525322 (NID:g12788815) "  
690. 1350  
/note="similar to Homo sapiens EST AL546207 (NID:g12879099) "  
719. 1190  
/note="similar to Homo sapiens EST AU574646 (NID:g12935046) "  
724. 1130  
/note="similar to Homo sapiens EST AA283104 (NID:g1926029) zt16a12.r1"  
967. 968  
/note="similar to Homo sapiens EST N64351 (NID:g1212180) yz47e12.s1"  
1007. 1175  
/note="similar to Homo sapiens EST N64351 (NID:g1212180) yz47e12.s1"  
1017. 1362  
/note="match to EST AL036737 (NID:g5927885) "  
1017. 1362  
/note="similar to Homo sapiens EST ALJ35195 (NID:g6603382) "  
1074. 1350  
/note="similar to EST BM671395 (NID:g18981293) "  
1135. 1681  
/note="similar to EST BM678317 (NID:g18988213) "  
1145  
/note="similar to Homo sapiens EST AI741715 (NID:g5110003) wg22b05.x1"  
1206. 1356  
/note="similar to Homo sapiens EST AU525280 (NID:g12788773) "  
1350. 1387  
/rpt\_family="(CA)n"  
1380. 1498  
/note="similar to Homo sapiens EST AU525280 (NID:g12788773) "  
1385. 1603  
/note="match to EST BG675536 (NID:g13906932) "  
1386. 1595  
/note="match to EST AL036737 (NID:g5927885) "  
1386. 1561  
/note="similar to Homo sapiens EST AU525322 (NID:g12788815) "  
1386. 1438  
/note="similar to Homo sapiens EST ALJ35195 (NID:g6603382) "  
1396. 1595  
/note="match to EST AW515614 (NID:g7153696) hd77d06.x1"  
1397. 1981  
/note="match to EST AM007714 (NID:g5856492) wt68c03.x1"

```

misc_feature 1403..1595
/Note="match to EST A1654124 (NID:g4738103) ty62a06.x1"
misc_feature 1404..1595
/Note="similar to Mus musculus EST BF152508
(NID:g11033903) uz34b1.y1"
misc_feature 1414..1716
/Note="similar to Homo sapiens EST AA406045 (NID:g2064073)
zu67d07.s1"
misc_feature 1422..1841
/Note="match to EST AA292874 (NID:g1941855) zt66a07.r1"
misc_feature 1438..2084
/Note="similar to Mus musculus EST AV258065
(NID:g16388469)"
misc_feature 1472..2203

```

```

Query Match 1.2%; Score 20; DB 9; Length 48546;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 19 AATGACAAATTAAGTATT 38
|||||
Db 12988 AATGACAAATTAAGTATT 13007

```

```

RESULT 35
ALJ34721/c 57070 bp DNA linear PRI 15-NOV-2001
DEFINITION Human DNA sequence from clone RP11-373E22 on chromosome 1, complete
sequence.
ACCESSION ALJ34721
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 57070)
REFERENCE 1 (bases 1 to 57070)
AUTHORS Mashreghi-Mohammadi, M.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiry:
humquerry@sanger.ac.uk
Clonerequest@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:16030110.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats: all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; SW.,
SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-373E22 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6

```

IMPORTANT: This sequence is not the entire insert of clone RP11-373E22. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-373E22 is at 57070 in this

```

FEATURES
Source location/Qualifiers
1..57070
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-373E22"
/clone_lib="RPCI-11.2"
BASE COUNT 19390 a 10751 c 10316 g 16613 t
ORIGIN

```

```

Query Match 1.2%; Score 20; DB 9; Length 57070;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 AATGACAAATTAAGTATT 22
|||||
Db 35649 AATGACAAATTAAGTATT 35630

```

```

RESULT 36
AC113541 58810 bp DNA linear HTG 15-MAY-2002
LOCUS AC113541
DEFINITION Mus musculus clone RP23-185N2, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC113541
VERSION AC113541.2 GI:20800304
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 58810)
REFERENCE 1 (bases 1 to 58810)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-185N2
JOURNAL Unpublished
2 (bases 1 to 58810)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Mcdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneis, L., Mihova, T.,
Mingra, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C., H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rossetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

```

```

TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 58810)
REFERENCE 3 (bases 1 to 58810)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalil, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardy, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Huine, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,
Lander, E., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Mcdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneis, L., Mihova, T.,
Mingra, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C., H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rossetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

```

sequence. The true left end of clone RP11-18005 is at 28981 in this sequence. The true right end of clone RP4-677H15 is at 2000 in this sequence.

TITLE  
JOURNAL  
COMMENT

Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margus, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K., Meldrum, J., Meneses, L., Milnova, T., Milenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Noddu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Rella, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Rhmann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

Submitted (15-May-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 15, 2002 this sequence version replaced g1:19033725.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L24218

Center clone name: 185\_N\_2

NOTE: This record contains 72 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
\* 738 837: gap of 100 bp  
\* 838 1574: contig of 737 bp in length  
\* 1575 1674: gap of 100 bp  
\* 1675 2407: contig of 733 bp in length  
\* 2408 2507: gap of 100 bp  
\* 2508 3219: contig of 712 bp in length  
\* 3220 3319: gap of 100 bp  
\* 3320 4016: contig of 697 bp in length  
\* 4017 4116: gap of 100 bp  
\* 4117 4828: contig of 712 bp in length  
\* 4829 4928: gap of 100 bp  
\* 4929 5645: contig of 717 bp in length  
\* 5646 5745: gap of 100 bp  
\* 5746 6454: contig of 709 bp in length  
\* 6455 6554: gap of 100 bp  
\* 6555 7287: contig of 733 bp in length  
\* 7288 7387: gap of 100 bp  
\* 7388 8122: contig of 735 bp in length  
\* 8123 8222: gap of 100 bp  
\* 8223 8955: contig of 733 bp in length  
\* 8956 9055: gap of 100 bp  
\* 9056 9765: contig of 710 bp in length  
\* 9766 9865: gap of 100 bp  
\* 9866 10570: contig of 705 bp in length  
\* 10571 10670: gap of 100 bp  
\* 10671 11393: contig of 723 bp in length  
\* 11394 11493: gap of 100 bp  
\* 11494 12213: contig of 720 bp in length  
\* 12214 12313: gap of 100 bp  
\* 12314 13018: contig of 705 bp in length  
\* 13019 13118: gap of 100 bp  
\* 13119 13826: contig of 708 bp in length

13827 13926: gap of 100 bp  
13927 14556: contig of 730 bp in length  
14557 14756: gap of 100 bp  
14757 15465: contig of 709 bp in length  
15466 15565: gap of 100 bp  
15566 16301: contig of 736 bp in length  
16302 16401: gap of 100 bp  
16402 17137: contig of 736 bp in length  
17138 17237: gap of 100 bp  
17238 17950: contig of 713 bp in length  
17951 18050: gap of 100 bp  
18051 18772: contig of 722 bp in length  
18773 18872: gap of 100 bp  
18873 19569: contig of 697 bp in length  
19570 19669: gap of 100 bp  
19670 20409: contig of 740 bp in length  
20410 20509: gap of 100 bp  
20510 21216: contig of 707 bp in length  
21217 21316: gap of 100 bp  
21317 22044: contig of 728 bp in length  
22045 22144: gap of 100 bp  
22145 22882: contig of 738 bp in length  
22883 22982: gap of 100 bp  
22983 23712: contig of 730 bp in length  
23713 23812: gap of 100 bp  
23813 24530: contig of 718 bp in length  
24531 24630: gap of 100 bp  
24631 25349: contig of 719 bp in length  
25350 25449: gap of 100 bp  
25450 26166: contig of 717 bp in length  
26167 26266: gap of 100 bp  
26267 26976: contig of 710 bp in length  
26977 27076: gap of 100 bp  
27077 27803: contig of 727 bp in length  
27804 27903: gap of 100 bp  
27904 28613: contig of 710 bp in length  
28614 28713: gap of 100 bp  
28714 29416: contig of 703 bp in length  
29417 29516: gap of 100 bp  
29517 30245: contig of 729 bp in length  
30246 30345: gap of 100 bp  
30346 31088: contig of 743 bp in length  
31089 31188: gap of 100 bp  
31189 31923: contig of 735 bp in length  
31924 32023: gap of 100 bp  
32024 32729: contig of 706 bp in length  
32730 32829: gap of 100 bp  
32830 33538: contig of 709 bp in length  
33539 33638: gap of 100 bp  
33639 34345: contig of 707 bp in length  
34346 34445: gap of 100 bp  
34446 35161: contig of 716 bp in length  
35162 35261: gap of 100 bp  
35262 36000: contig of 739 bp in length  
36001 36100: gap of 100 bp  
36101 36805: contig of 705 bp in length  
36806 36905: gap of 100 bp  
36906 37643: contig of 738 bp in length  
37644 37743: gap of 100 bp  
37744 38479: contig of 736 bp in length  
38480 38579: gap of 100 bp  
38580 39303: contig of 724 bp in length  
39304 39403: gap of 100 bp  
39404 40130: contig of 727 bp in length  
40131 40230: gap of 100 bp  
40231 40908: contig of 678 bp in length  
40909 41008: gap of 100 bp  
41009 41728: contig of 720 bp in length  
41729 41828: gap of 100 bp  
41829 42534: contig of 706 bp in length  
42535 42634: gap of 100 bp  
42635 43333: contig of 699 bp in length  
43334 43433: gap of 100 bp



\* 16371 17070: contig of 700 bp in length  
\* 17071 17170: gap of 100 bp  
\* 17171 17886: contig of 716 bp in length  
\* 17887 17986: gap of 100 bp  
\* 17987 18696: contig of 710 bp in length  
\* 18697 18796: gap of 100 bp  
\* 18797 19516: contig of 720 bp in length  
\* 19517 19616: gap of 100 bp  
\* 19617 20334: contig of 718 bp in length  
\* 20335 20434: gap of 100 bp  
\* 20435 21160: contig of 726 bp in length  
\* 21161 21260: gap of 100 bp  
\* 21261 21993: contig of 733 bp in length  
\* 21994 22093: gap of 100 bp  
\* 22094 22840: contig of 747 bp in length  
\* 22841 22940: gap of 100 bp  
\* 22941 23664: contig of 724 bp in length  
\* 23665 23764: gap of 100 bp  
\* 23765 24448: contig of 684 bp in length  
\* 24449 24548: gap of 100 bp  
\* 24549 25266: contig of 718 bp in length  
\* 25267 25366: gap of 100 bp  
\* 25367 26078: contig of 712 bp in length  
\* 26079 26178: gap of 100 bp  
\* 26179 26891: contig of 713 bp in length  
\* 26892 26991: gap of 100 bp  
\* 26992 27684: contig of 693 bp in length  
\* 27685 27784: gap of 100 bp  
\* 27785 28489: contig of 705 bp in length  
\* 28490 28589: gap of 100 bp  
\* 28590 29303: contig of 714 bp in length  
\* 29304 29403: gap of 100 bp  
\* 29404 30139: contig of 736 bp in length  
\* 30140 30239: gap of 100 bp  
\* 30240 30960: contig of 721 bp in length  
\* 30961 31060: gap of 100 bp  
\* 31061 31764: contig of 704 bp in length  
\* 31765 31864: gap of 100 bp  
\* 31865 32583: contig of 719 bp in length  
\* 32584 32683: gap of 100 bp  
\* 32684 33394: contig of 711 bp in length  
\* 33395 33494: gap of 100 bp  
\* 33495 34213: contig of 719 bp in length  
\* 34214 34313: gap of 100 bp  
\* 34314 35030: contig of 717 bp in length  
\* 35031 35130: gap of 100 bp  
\* 35131 35844: contig of 714 bp in length  
\* 35845 35944: gap of 100 bp  
\* 35945 36669: contig of 725 bp in length  
\* 36670 36769: gap of 100 bp  
\* 36770 37495: contig of 726 bp in length  
\* 37496 37595: gap of 100 bp  
\* 37596 38320: contig of 725 bp in length  
\* 38321 38420: gap of 100 bp  
\* 38421 39162: contig of 742 bp in length  
\* 39163 39262: gap of 100 bp  
\* 39263 40009: contig of 747 bp in length  
\* 40010 40109: gap of 100 bp  
\* 40110 40821: contig of 712 bp in length  
\* 40822 40921: gap of 100 bp  
\* 40922 41627: contig of 706 bp in length  
\* 41628 41727: gap of 100 bp  
\* 41728 42437: contig of 710 bp in length  
\* 42438 42537: gap of 100 bp  
\* 42538 43250: contig of 713 bp in length  
\* 43251 43350: gap of 100 bp  
\* 43351 44066: contig of 716 bp in length  
\* 44067 44166: gap of 100 bp  
\* 44167 44882: contig of 716 bp in length  
\* 44883 44982: gap of 100 bp  
\* 44983 45702: contig of 720 bp in length

Query Match 1.2% Score 20: DB 2: Length 65259;

Best Local Similarity 100.0%: Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 622 GAGCTAATATTCCTTACTA 641  
|||||  
Db 25639 GAGCTAATATTCCTTACTA 25658

RESULT 38  
AC114147/C  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-144p3, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 45 unordered pieces.  
AC114147.2 GI:21732759  
AC114147  
KEYWORDS  
HTG: HTGS\_PHASE1.  
SOURCE  
Rattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
1 (bases 1 to 74010)  
Muzny,D.M., Adams,C., Adjo-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alspbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonin,D.,  
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,J., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,  
Hornst,E., Howard,S., Huber,J., Huliyk,S., Hume,D., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlssoon,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtharge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Matinez,E.,  
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogunu,M., Okunonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peety,J., Perez,L.,  
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Saverly,G.,  
Scherer,S., Scott,G., Shen,H., Shoshart,N., Sisson,L.,  
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansley,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,  
Umani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 74010)  
Worley,K.C.  
Direct Submission  
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 74010)  
Worley,K.C.  
Direct Submission

## JOURNAL

## COMMENT

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2002 this sequence version replaced g1:19224474.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GLOB

Center clone name: CH230-144F3

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 24661 bases at least Q40

Consensus quality: 26589 bases at least Q30

Consensus quality: 27921 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1      1381: contig of 1381 bp in length
1382      1481: gap of unknown length
1482      2791: contig of 1310 bp in length
2792      2891: gap of unknown length
2892      4184: contig of 1293 bp in length
4185      4284: gap of unknown length
4285      5398: contig of 1114 bp in length
5399      5498: gap of unknown length
5499      6596: contig of 1098 bp in length
6597      6696: gap of unknown length
6697      7726: contig of 1030 bp in length
7727      7826: gap of unknown length
7827      8916: contig of 1090 bp in length
8917      9016: gap of unknown length
9017      10122: contig of 1106 bp in length
10123      11222: gap of unknown length
11223      11326: contig of 1004 bp in length
11327      12773: contig of 1447 bp in length
12774      12873: gap of unknown length
12873      13949: contig of 1076 bp in length
13950      14049: gap of unknown length
14050      15641: contig of 1592 bp in length
15642      15741: gap of unknown length
15742      17427: contig of 1686 bp in length
17428      17527: gap of unknown length
17528      18919: contig of 1392 bp in length
18920      19019: gap of unknown length
19020      20050: contig of 1031 bp in length
20051      21271: contig of 1121 bp in length
21272      21371: gap of unknown length
21372      22617: contig of 1246 bp in length
22618      22717: gap of unknown length
22718      23733: contig of 1016 bp in length
23734      23833: gap of unknown length
23834      25251: contig of 1418 bp in length
25252      25351: gap of unknown length
25352      26502: contig of 1151 bp in length
26503      28122: contig of 1520 bp in length
28123      28222: gap of unknown length
28223      29582: contig of 1360 bp in length

```

```

*      29583      29682: gap of unknown length
*      29683      31194: contig of 1512 bp in length
*      31195      31294: gap of unknown length
*      31295      32524: contig of 1230 bp in length
*      32525      32624: gap of unknown length
*      32625      34487: contig of 1863 bp in length
*      34488      34587: gap of unknown length
*      34588      36096: contig of 1509 bp in length
*      36097      37881: gap of unknown length
*      37882      37981: gap of unknown length
*      37982      39421: contig of 1440 bp in length
*      39422      39521: gap of unknown length
*      39522      41762: contig of 2241 bp in length
*      41763      41862: gap of unknown length
*      41863      43073: contig of 1211 bp in length
*      43074      43173: gap of unknown length
*      43174      44522: contig of 1349 bp in length
*      44523      44622: gap of unknown length
*      44623      46552: contig of 1930 bp in length
*      46553      48103: gap of unknown length
*      48104      48203: contig of 1451 bp in length
*      48204      49906: gap of unknown length
*      49907      50006: contig of 1703 bp in length
*      50007      51069: contig of 1063 bp in length
*      51070      51169: gap of unknown length
*      51170      52193: contig of 1024 bp in length
*      52194      52293: gap of unknown length
*      52294      53689: contig of 1386 bp in length
*      53690      53789: gap of unknown length
*      53790      55812: contig of 2023 bp in length
*      55813      55912: gap of unknown length
*      55913      58145: contig of 2233 bp in length
*      58146      58245: gap of unknown length
*      58246      59827: contig of 1682 bp in length
*      59828      60027: gap of unknown length
*      60028      61826: contig of 1799 bp in length
*      61827      61926: gap of unknown length
*      61927      64162: contig of 2236 bp in length
*      64163      64262: gap of unknown length
*      64263      67752: contig of 3490 bp in length
*      67753      67852: gap of unknown length
*      67853      71791: contig of 3939 bp in length
*      71792      71891: gap of unknown length
*      71892      74010: contig of 2119 bp in length.

```

FEATURES

source 1. 74010 Location/Qualifiers

BASE COUNT 17870 a 15598 c 17993 g 18093 t 4456 others

ORIGIN

Query Match 1.2%; Score 20; DB 2; Length 74010;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1482 TTCATCATTCGATTTGGTT 1501  
|||||

Db 73928 TTCATCATTCGATTTGGTT 73909

RESULT 39

AC087717 74274 bp DNA linear HTG 16-JAN-2001

LOCUS AC087717

DEFINITION Homo sapiens chromosome 11 clone RPI-145W24 map 11, LOW-PASS

SEQUENCE SAMPLING.

ACCESSION AC087717

VERSION AC087717.1 GI:12232529

KEYWORDS HTG: HTGS\_PHASE0.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 74274)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 11, clone RP1-145M24  
Unpublished

2 (bases 1 to 74274)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barra,N., Bastien,Y., Boguski,M., Bouckhalter,B., Brown,A.,  
Camarata,J., Campilano,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,  
Gardina,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heathford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,  
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,  
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,  
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,  
Roy,A., Santos,R., Schauer,B., Stange-Thomann,N., Stojanovic,N.,  
Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,  
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Submitted (16-JAN-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L12105

Center clone name: 145\_M24

\* NOTE: This record contains 89 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 750: contig of 750 bp in length  
\* 751 850: gap of 100 bp  
\* 851 1637: contig of 787 bp in length  
\* 1638 1737: gap of 100 bp  
\* 1738 2478: contig of 741 bp in length  
\* 2479 2578: gap of 100 bp  
\* 2579 3294: contig of 716 bp in length  
\* 3295 3394: gap of 100 bp  
\* 3395 4106: contig of 712 bp in length  
\* 4107 4206: gap of 100 bp  
\* 4207 4886: contig of 680 bp in length  
\* 4887 4986: gap of 100 bp  
\* 4987 5679: contig of 693 bp in length  
\* 5680 5779: gap of 100 bp  
\* 5780 6507: contig of 728 bp in length  
\* 6508 6607: gap of 100 bp  
\* 6608 7351: contig of 744 bp in length  
\* 7352 7451: gap of 100 bp  
\* 7452 8211: contig of 760 bp in length

\* 8212 8311: gap of 100 bp  
\* 8312 9026: contig of 715 bp in length  
\* 9027 9126: gap of 100 bp  
\* 9127 9881: contig of 755 bp in length  
\* 9882 9981: gap of 100 bp  
\* 9982 10716: contig of 735 bp in length  
\* 10717 10816: gap of 100 bp  
\* 10817 11535: contig of 719 bp in length  
\* 11536 11635: gap of 100 bp  
\* 11636 12299: contig of 664 bp in length  
\* 12300 12399: gap of 100 bp  
\* 12400 13160: contig of 761 bp in length  
\* 13161 13260: gap of 100 bp  
\* 13261 13998: contig of 738 bp in length  
\* 13999 14098: gap of 100 bp  
\* 14099 14831: contig of 733 bp in length  
\* 14832 14931: gap of 100 bp  
\* 14932 15680: contig of 749 bp in length  
\* 15681 15780: gap of 100 bp  
\* 15781 16526: contig of 746 bp in length  
\* 16527 16626: gap of 100 bp  
\* 16627 17394: contig of 768 bp in length  
\* 17395 17494: gap of 100 bp  
\* 17495 18212: contig of 718 bp in length  
\* 18213 18312: gap of 100 bp  
\* 18313 19052: contig of 740 bp in length  
\* 19053 19152: gap of 100 bp  
\* 19153 19912: contig of 760 bp in length  
\* 19913 20012: gap of 100 bp  
\* 20013 20740: contig of 728 bp in length  
\* 20741 20840: gap of 100 bp  
\* 20841 21566: contig of 726 bp in length  
\* 21567 21666: gap of 100 bp  
\* 21667 22334: contig of 668 bp in length  
\* 22335 22434: gap of 100 bp  
\* 22435 23181: contig of 747 bp in length  
\* 23182 23281: gap of 100 bp  
\* 23282 24020: contig of 739 bp in length  
\* 24021 24120: gap of 100 bp  
\* 24121 24855: contig of 735 bp in length  
\* 24856 24955: gap of 100 bp  
\* 24956 25709: contig of 754 bp in length  
\* 25710 25809: gap of 100 bp  
\* 25810 26550: contig of 741 bp in length  
\* 26551 26650: gap of 100 bp  
\* 26651 27392: contig of 742 bp in length  
\* 27393 27492: gap of 100 bp  
\* 27493 28232: contig of 740 bp in length  
\* 28233 28332: gap of 100 bp  
\* 28333 29064: contig of 732 bp in length  
\* 29065 29164: gap of 100 bp  
\* 29165 29884: contig of 720 bp in length  
\* 29885 29984: gap of 100 bp  
\* 29985 30676: contig of 692 bp in length  
\* 30677 30776: gap of 100 bp  
\* 30777 31527: contig of 751 bp in length  
\* 31528 31627: gap of 100 bp  
\* 31628 32362: contig of 735 bp in length  
\* 32363 32462: gap of 100 bp  
\* 32463 33210: contig of 748 bp in length  
\* 33211 33310: gap of 100 bp  
\* 33311 34076: contig of 766 bp in length  
\* 34077 34176: gap of 100 bp  
\* 34177 34935: contig of 759 bp in length  
\* 34936 35035: gap of 100 bp  
\* 35036 35756: contig of 721 bp in length  
\* 35757 35856: gap of 100 bp  
\* 35857 36585: contig of 729 bp in length  
\* 36586 36685: gap of 100 bp  
\* 36686 37431: contig of 746 bp in length  
\* 37432 37531: gap of 100 bp  
\* 37532 38275: contig of 744 bp in length  
\* 38276 38375: gap of 100 bp

```

* 38376 39094: contig of 719 bp in length
* 39095 39194: gap of 100 bp
* 39195 39960: contig of 766 bp in length
* 39961 40060: gap of 100 bp
* 40061 40779: contig of 719 bp in length
* 40780 40879: gap of 100 bp
* 40880 41612: contig of 733 bp in length
* 41613 41712: gap of 100 bp
* 41713 42441: contig of 729 bp in length
* 42442 42541: gap of 100 bp
* 42542 43275: contig of 734 bp in length
* 43276 43375: gap of 100 bp
* 43376 44114: contig of 739 bp in length
* 44115 44214: gap of 100 bp
* 44215 44965: contig of 751 bp in length
* 44966 45065: gap of 100 bp
* 45066 45789: contig of 724 bp in length
* 45790 45889: gap of 100 bp
* 45890 46633: contig of 744 bp in length
* 46634 46733: gap of 100 bp
* 46734 47462: contig of 729 bp in length
* 47463 47562: gap of 100 bp
* 47563 48301: contig of 739 bp in length
* 48302 48401: gap of 100 bp
* 48402 49125: contig of 724 bp in length
* 49126 49225: gap of 100 bp
* 49226 49962: contig of 737 bp in length
* 49963 50062: gap of 100 bp
* 50063 50798: contig of 736 bp in length
* 50799 50898: gap of 100 bp
* 50899 51635: contig of 737 bp in length
* 51636 51735: gap of 100 bp
* 51736 52464: contig of 729 bp in length
* 52465 52564: gap of 100 bp
* 52565 53316: contig of 753 bp in length
* 53317 53416: gap of 100 bp
* 53417 54176: contig of 760 bp in length
* 54177 54276: gap of 100 bp
* 54277 55004: contig of 728 bp in length
* 55005 55104: gap of 100 bp
* 55105 55867: contig of 763 bp in length
* 55868 55967: gap of 100 bp
* 55968 56676: contig of 709 bp in length
* 56677 56776: gap of 100 bp
* 56777 57511: contig of 737 bp in length
* 57512 57611: gap of 100 bp

```

Query Match 1.2%: Score 20; DB 2; Length 74274;  
 Best Local Similarity 100.0%; Pred. No. 36;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 187 ATCAAAAGACACTCTTATT 206  
 ||||||||||||||||  
 Db 43542 ATCAAAAGACACTCTTATT 43561

RESULT 40  
 AC025723 77422 bp DNA linear INV 29-MAY-2002  
 LOCUS Caenorhabditis elegans cosmid Y54F10AM, complete sequence.  
 AC025723  
 AC025723.2 GI:13559726  
 VERSION  
 KEYWORDS  
 SOURCE HTG.  
 ORGANISM  
 Caenorhabditis elegans.  
 Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 77422)  
 Waterston, R.  
 Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium  
 JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 99069613

PUBMED 9851916  
 REFERENCE 2 (bases 1 to 77422)  
 AUTHORS Ryan, E., Courtney, L. and Yoakum, M.  
 TITLE The sequence of C. elegans cosmid Y54F10AM  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 77422)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 77422)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-APR-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 5 (bases 1 to 77422)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
 REFERENCE 6 (bases 1 to 77422)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
 COMMENT On Apr 6, 2001 this sequence version replaced gi:7230347.  
 Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=Y54F10AM;class=Sequence>

#### NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y54F10AL, 6000 bp overlap; the 3' cosmid is Y54F10AR, 200 bp overlap.

#### NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yui Kohara ([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C. elegans ORFeome cloning project (<http://worfdb.dicli.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 20:35:19 ; Search time 1670 Seconds  
(without alignments)  
15875.456 Million cell updates/sec

Title: US-09-869-433-1  
Perfect score: 1637  
Sequence: 1 gaataaaaaactacacagaa.....gttgatccctacgtctttc 1637

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

- EST:\*  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrc:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1.5	755	17	BH837547	BH837547 LMCRO5001
2	1.4	483	17	BH773218	BH773218 uzm0002f0
3	1.3	770	17	AQ893684	AQ893684 HS_3148_A
4	1.3	292	9	AA639627	AA639627 ng83d01.s
5	1.3	330	17	AQ013184	AQ013184 CIT-HSP-2
6	1.3	346	9	A1002102	A1002102 ct38d04.s

7	21	1.3	346	10	AM090585	AM090585 xc84903.x
8	21	1.3	420	14	B0012918	B0012918 UT-1-BC1P
9	21	1.3	509	9	AA644559	AA644559 ab63g08.s
10	21	1.3	597	17	A2865523	A2865523 2M0175K12
11	21	1.3	951	21	BH164112	BH164112 ENTB33TF
12	21	1.3	1017	14	BQ048904	BQ048904 AGENECOURT
13	21	1.2	119	17	AQ115988	AQ115988 RPCI11-57
14	20	1.2	254	14	BQ610731	BQ610731 sap50805.
15	20	1.2	289	10	BA836799	BA836799 BA36799
16	20	1.2	297	10	AM374096	AM374096 PM4-BT054
17	20	1.2	303	10	BB216685	BB216685 BB216685
18	20	1.2	312	10	BB381288	BB381288 BB381288
19	20	1.2	328	9	AA712844	AA712844 32403.Lam
20	20	1.2	334	17	AQ697744	AQ697744 HS_5553_A
21	20	1.2	337	17	AZ016529	AZ016529 RPCI-23-3
22	20	1.2	339	10	BE432018	BE432018 EST398547
23	20	1.2	342	17	AQ685833	AQ685833 HS_5549_A
24	20	1.2	352	12	BG628369	BG628369 cc-estfcl
25	20	1.2	356	12	BG626629	BG626629 cc-estfcl
26	20	1.2	368	12	BG628526	BG628526 cc-estfcl
27	20	1.2	388	10	AM011059	AM011059 SRT16C01.P
28	20	1.2	439	17	AZ630188	AZ630188 1M0483D12
29	20	1.2	439	10	AM218715	AM218715 EST301195
30	20	1.2	440	10	AM218716	AM218716 EST301196
31	20	1.2	458	10	AM933403	AM933403 EST359342
32	20	1.2	472	17	AQ433974	AQ433974 HS_5088_A
33	20	1.2	496	12	BF098187	BF098187 EST428796
34	20	1.2	498	10	AM651372	AM651372 EST329826
35	20	1.2	500	13	BM404442	BM404442 EST578769
36	20	1.2	514	10	AM221154	AM221154 EST297623
37	20	1.2	546	10	BB284364	BB284364 BB284364
38	20	1.2	548	17	AQ121410	AQ121410 HS_3087_B
39	20	1.2	558	14	BQ119131	BQ119131 EST604707
40	20	1.2	559	10	BE459902	BE459902 EST415194
41	20	1.2	566	10	AM907382	AM907382 EST343505
42	20	1.2	569	9	A1993855	A1993855 701515173
43	20	1.2	575	14	BQ116413	BQ116413 EST501989
44	20	1.2	576	13	BM112431	BM112431 EST559667
45	20	1.2	579	12	BE919839	BE919839 EST423608
46	20	1.2	580	10	BE41982	BE41982 EST413320
47	20	1.2	582	10	BE460070	BE460070 EST415362
48	20	1.2	585	10	AM035953	AM035953 EST282812
49	20	1.2	589	10	BE451607	BE451607 EST402483
50	20	1.2	594	17	AZ798455	AZ798455 2M0055117
51	20	1.2	595	10	AM930772	AM930772 EST356615
52	20	1.2	596	10	AV751415	AV751415 AV751415
53	20	1.2	596	10	AM221000	AM221000 EST297469
54	20	1.2	597	10	AM932225	AM932225 EST356068
55	20	1.2	604	17	AZ821248	AZ821248 2M0093122
56	20	1.2	619	9	A1771102	A1771102 EST252202
57	20	1.2	619	12	BG644061	BG644061 EST512235
58	20	1.2	623	14	BQ046082	BQ046082 EST595200
59	20	1.2	629	13	BM113917	BM113917 EST561453
60	20	1.2	633	12	BG589972	BG589972 EST497814
61	20	1.2	648	17	AZ625761	AZ625761 1M0465N04
62	20	1.2	657	10	AM773675	AM773675 EST332661
63	20	1.2	661	10	BB150648	BB150648 BB150648
64	20	1.2	666	17	AG053391	AG053391 Pan.trog1
65	20	1.2	671	12	BG594974	BG594974 EST493652
66	20	1.2	676	10	BE461454	BE461454 EST412873
67	20	1.2	685	10	AM040455	AM040455 EST283415
68	20	1.2	685	17	AZ338986	AZ338986 1M0070H16
69	20	1.2	690	13	BI922435	BI922435 EST542339
70	20	1.2	695	13	BI406811	BI406811 161B12.Ma
71	20	1.2	713	13	BI923357	BI923357 EST543416
72	20	1.2	719	13	BM410984	BM410984 EST585311
73	20	1.2	720	12	BG890033	BG890033 EST515884
74	20	1.2	725	12	BG598397	BG598397 EST503297
75	20	1.2	725	17	AQ247939	AQ247939 HS_2064_LB
76	20	1.2	732	12	BG125387	BG125387 EST471033
77	20	1.2	771	12	BF573280	BF573280 602079671
78	20	1.2	844	14	BQ510145	BQ510145 EST617560
79	20	1.2	850	17	AZ540960	AZ540960 ENTB56TR

```

80 20 1.2 888 17 A2546199
81 20 1.2 902 17 BH149116
82 20 1.2 1068 17 BH663206
83 19 1.2 195 14 BQ234495
84 19 1.2 283 12 BF516008
85 19 1.2 296 10 AM435908
86 19 1.2 304 17 A2790186
87 19 1.2 313 10 BB208477
88 19 1.2 353 17 AQ237639
89 19 1.2 359 17 AQ302835
90 19 1.2 360 9 AA591598
91 19 1.2 383 17 AQ145028
92 19 1.2 399 9 AA516865
93 19 1.2 401 10 AM815196
94 19 1.2 406 17 AQ047250
95 19 1.2 411 10 AM390055
96 19 1.2 414 10 BE090450
97 19 1.2 417 9 AA767956
98 19 1.2 418 10 BB686390
99 19 1.2 420 9 AJ462906
c 100 19 1.2 426 10 AM208986

```

## ALIGNMENTS

```

RESULT 1
BH837547/c 755 bp DNA linear GSS 28-MAY-2002
LOCUS LMKR050015H09f Zea mays L. Zea mays genomic clone LMKR050015H09f,
DEFINITION DNA sequence.
ACCESSION BH837547
VERSION BH837547.1 GI:21235425
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays.

```

## REFERENCE

```

AUTHORS Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
JOURNAL 1 (bases 1 to 755)
COMMENT Kim, S.W., Yu, Y., Lee, M.C., Main, D., and Wang, R.A.
Methyl-filtration genomic sequence from maize
Unpublished (2002)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total High Quality bases = 520
Seq primer: TAATPACGACCTCATATGCG
Class: shotgun

```

## FEATURES

```

FEATURES
source
1..755
Location/Qualifiers
/organism="Zea mays"
/strain="B73"
/db_xref="taxon:4577"
/cclone="LMKR050015H09f"
/cclone_lib="Zea mays L."
/tissue_type="leaf"
/lab_host="DH10B"
/note="Vector: pGEM-T easy; Site_1: Mcr BC;
Methyl-filtration library; Nuclei DNA was completely
digested with Mcr BC, size fractionated and transformed
to E.Coli DH10B."
High quality sequence start: 12
High quality sequence stop: 714.
Location/Qualifiers
1..755

```

## BASE COUNT

```

BASE COUNT 242 a 176 c 152 g 179 t 6 others
ORIGIN
Query Match 1.5%; Score 25; DB 17; Length 755;
Best Local Similarity 100.0%; Pred. No. 0.36;

```

```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 20 ATAGAAATAAAGTATTTCAGAG 44
Db 299 ATAGAAATAAAGTATTTCAGAG 275

```

```

RESULT 2
BH773218 483 bp DNA linear GSS 28-MAR-2002
LOCUS uzmb002f003c09 5', DNA sequence.
DEFINITION uzmb002f003c09f uzmb unfiltered library Zea mays genomic clone
uzmb002f003c09 5', DNA sequence.
ACCESSION BH773218
VERSION BH773218.1 GI:19775208
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays.

```

```

REFERENCE
AUTHORS Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
JOURNAL 1 (bases 1 to 483)
COMMENT Budiman, M.A., Freese, R.G., Bedell, J.A., Nunberg, A.N. and Lakey, N.D.
Genethresher methylation filtered genomic sequences from maize
Unpublished (2002)
Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 5975
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: uzmb002f003 row: c column: 09
Seq primer: M13 forward
Class: shotgun

```

## FEATURES

```

FEATURES
source
1..483
Location/Qualifiers
/organism="Zea mays"
/cultivar="MO17"
/db_xref="taxon:4577"
/cclone="uzmb002f003c09"
/cclone_lib="uzmb unfiltered library"
/note="Organ: leaf; Vector: pGEM-T(-); Site_1: HincII; DNA
prepared from purified nuclei was randomly sheared,
end-repaired, size fractionated to enrich for the 0.5 to
5 kb fraction, ligated into HincII-digested pGEM-T(-)
vector and electroporated into E. coli cells."
High quality sequence stop: 483.
Location/Qualifiers
1..483

```

## BASE COUNT

```

BASE COUNT 128 a 98 c 93 g 164 t
ORIGIN
Query Match 1.4%; Score 23; DB 17; Length 483;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 20 ATAGAAATAAAGTATTTCAGA 42
Db 428 ATAGAAATAAAGTATTTCAGA 450

```

```

RESULT 3
AO893684/c 770 bp DNA linear GSS 10-NOV-1999
LOCUS HS_3148_A2.C11.MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-3148 COL=22 Row=E, DNA sequence.
ACCESSION AO893684
VERSION AO893684.1 GI:6349874
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and

```

```

TITLE
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE
99380589
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
BAC end Web Server: http://www.husc.washington.edu
Plate: 3148 row: E Column: 22
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 770.
FEATURES
source
1..770
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3148 Col=22 Row=E"
/clone_11b="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT
245 a 144 c 108 g 273 t
ORIGIN
Query Match
Best Local Similarity 1.3%; Score 22; DB 17; Length 770;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 AAGAAAAACCTTTGGAAAAATT 85
|||||
Db 702 AAGAAAAACCTTTGGAAAAATT 681

RESULT 4
AA639627 292 bp mRNA linear EST 23-OCT-1997
ng83d01.sl NCI-CGAP_C09 Homo sapiens cDNA IMAGE:1158913 3',
LOCUS
mRNA sequence.
ACCESSION
AA639627
VERSION
AA639627.1 GI:2563406
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 292)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
www.bio.lnsl.gov/bbrp/image/image.html
Seq primer: -40m3 fwd. RT from Amersham
High quality sequence stop: 282.
FEATURES
source
1..292
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1158913"

```

```

/clone_11b="NCI CGAP_C09"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo (Soares4
).
BASE COUNT
109 a 63 c 37 g 83 t
ORIGIN
Query Match
Best Local Similarity 1.3%; Score 21; DB 9; Length 292;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 TAGATCCTGGGGAATTCTT 720
|||||
Db 289 TAGATCCTGGGGAATTCTT 269

RESULT 5
AA013184/c 330 bp DNA linear GSS 06-JUN-1998
LOCUS
CIT-HSP-2298E18.TR CIT-HSP Homo sapiens genomic clone 2298E18, DNA
DEFINITION
sequence.
ACCESSION
AA013184
VERSION
AA013184.1 GI:3185749
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 330)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
JOURNAL
Other_GSSS: CIT-HSP-2298E18.TF
COMMENT
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
source
1..330
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7153678"
/db_xref="taxon:9606"
/clone="2298E18"
/clone_11b="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT
112 a 58 c 67 g 93 t
ORIGIN
Query Match
Best Local Similarity 1.3%; Score 21; DB 17; Length 330;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 187 ATACAAAGACACTTATTTG 207  
LOCUS 142 ATACAAAGACACTTATTTG 122

RESULT 6  
AI002102/c 346 bp mRNA linear EST 27-AUG-1998  
LOCUS 0136d04.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1619047  
DEFINITION 3', mRNA sequence.

ACCESSION AI002102  
VERSION AI002102.1 GI:3202139  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 346)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
, Ph.D.

FEATURES  
source  
1. 346  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1619047"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT7T3-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech Laboratories  
, Inc., and primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAAGTGGAGCGCCGCCCAATTTTATTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization to Col5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 125 a 70 c 40 g 111 t

ORIGIN

Query Match 1.3%; Score 21; DB 9; Length 346;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 TAGATCTTGGGGAATTTCTT 720  
LOCUS 303 TAGATCTTGGGGAATTTCTT 283

RESULT 7  
AM090585/c 346 bp mRNA linear EST 15-OCT-1999  
LOCUS xc84903.x1 NCI-CGAP\_Brn35 Homo sapiens cDNA clone IMAGE:259096 3',  
DEFINITION mRNA sequence.  
ACCESSION AM090585  
VERSION AM090585.1 GI:6047929  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 346)  
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/RTGAP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

FEATURES  
source  
1. 346  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:259096"  
/clone\_lib="NCI-CGAP\_Brn35"  
/tissue\_type="tumor, 5 pooled (see description)"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-Sport6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: oligo dr.  
Average insert size 1.33 kb. Tumor types include:  
meningioma, oligodendroglioma, astrocytoma (grade II),  
medulloblastoma, astrocytoma (grade IV). Life Technologies  
catalog # : 11544-012"

BASE COUNT 124 a 69 c 40 g 113 t

ORIGIN

Query Match 1.3%; Score 21; DB 10; Length 346;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 TAGATCTTGGGGAATTTCTT 720  
LOCUS 312 TAGATCTTGGGGAATTTCTT 292

RESULT 8  
B0012918/c 420 bp mRNA linear EST 26-MAR-2002  
LOCUS UT-1-BC1P-ava-e-07-0-UT.s1 NCI-CGAP\_P13 Homo sapiens cDNA clone  
DEFINITION UT-1-BC1P-ava-e-07-0-UT 3', mRNA sequence.

ACCESSION B0012918  
VERSION B0012918.1 GI:19737819  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 420)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Steven Brown  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@iowa.edu  
Seq primer: M13 FORWARD  
POLYA=Yes

# FEATURES

Source

1. 420  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-1-BC1p-ava-e-07-0-UI"  
/clone\_lib="NCLCGAP.pl3"  
/tissue\_type="Placenta"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Placenta; Vector: pT73-Pac (pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I; NCLCGAP.pl3 is a subcloned cDNA library constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are GA, AGGA. For additional information, contact: Bento Soares, bento-soares@iowa.edu  
TAG\_Lib=UI-1-BC1p  
TAG\_ISSUE=placenta human 8 week  
TAG\_SEQ=GA"

BASE COUNT 143 a 83 c 46 g 148 t  
ORIGIN

Query Match 1.3%; Score 21; DB 14; Length 420;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 TAGATCCTGGGGAATTCTT 720  
|||||  
Db 308 TAGATCCTGGGGAATTCTT 288

RESULT 9  
AA644559/c 509 bp mRNA linear EST 06-MAR-1998  
LOCUS  
DEFINITION  
IMAGE:84534 3', mRNA sequence.  
ACCESSION  
AA644559  
VERSION  
AA644559.1 GI:2569777  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 509)  
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Merra,M., Martin,J., Moore,B., Schellberg,K., Stepcoe,M., Tan,F., Theisling,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1206 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 463.  
Location/Qualifiers

# FEATURES

Source

1. 509  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:845534"  
/clone\_lib="Stratagene lung carcinoma 937218"  
/tissue\_type="lung Carcinoma"  
/cell\_line="NCI-H69"  
/dev\_stage="cell line NCI-H69"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Small cell carcinoma cell line NCI-H69. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGACGACG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3"

BASE COUNT 173 a 107 c 67 g 162 t  
ORIGIN

Query Match 1.3%; Score 21; DB 9; Length 509;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 TAGATCCTGGGGAATTCTT 720  
|||||  
Db 292 TAGATCCTGGGGAATTCTT 272

RESULT 10  
A2865523 597 bp DNA linear GSS 21-FEB-2001  
LOCUS  
DEFINITION  
2M0175K12P Mouse 10kb plasmid U0GC1M library Mus musculus genomic  
clone U0GC2M0175K12 R, DNA sequence.  
ACCESSION  
A2865523  
VERSION  
A2865523.1 GI:13065920  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Scurionahli; Muridae; Murineae; Mus.  
REFERENCE  
1 (bases 1 to 597)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0175 row: K column: 12  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 597.  
Location/Qualifiers

# FEATURES

Source

1. 597  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U0GC2M0175K12"  
/clone\_lib="Mouse 10kb plasmid U0GC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.3 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (914732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 194 a 148 c 89 g 166 t

Query Match 1.3%; Score 21; DB 17; Length 597;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 873 AAGATGAGCTTCCTATCTT 893  
|||||  
DB 184 AAGATGAGCTTCCTATCTT 204

## RESULT 11

BH164112 951 bp DNA Linear GSS 24-SEP-2001  
LOCUS ENT1B33TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic, DNA sequence.

ACCESSION BH164112  
VERSION BH164112.1 GI:15737550  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica  
REFERENCE 1 (bases 1 to 951)  
AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library (2001)

JOURNAL Unpublished (2001)  
COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
Seq primer: M13-Forward  
Class: Shotgun  
High quality sequence start: 15  
High quality sequence stop: 754.  
Location/Qualifiers

FEATURES  
source 1..951  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_id="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHD10B (phage-resistant)"  
Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 452 a 86 c 132 g 281 t

OY 18 GAATGAAATATAAGTATTT 38  
|||||  
DB 717 GAATGAAATATAAGTATTT 737

## RESULT 12

BQ048904 1017 bp mRNA Linear EST 29-MAR-2002  
LOCUS AGENCOURT\_6832569 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:5789594  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ048904  
VERSION BQ048904.1 GI:19808244  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1017)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs@email.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM12883 row: 9 column: 03  
High quality sequence stop: 699.  
Location/Qualifiers

FEATURES  
source 1..1017  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="IMAGE:5789594"  
/clone\_id="NIH\_MGC\_92"  
/issue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."

BASE COUNT 339 a 195 c 219 g 264 t

Query Match 1.3%; Score 21; DB 14; Length 1017;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 AGAANAACCTTTGGAAATTT 85  
|||||  
DB 985 AGAANAACCTTTGGAAATTT 1005

## RESULT 13

AQ115988 119 bp DNA Linear GSS 20-APR-1999  
LOCUS RC111-57F1.TK.1 RPT-11 Homo sapiens genomic clone RPT-11-57F1.  
DEFINITION DNA sequence.  
ACCESSION AQ115988  
VERSION AQ115988.1 GI:3492109



**KEYWORDS** GSS.  
**SOURCE** human.  
**ORGANISM** Homo sapiens

**REFERENCE** 1  
**AUTHORS** Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Liner,K., Golden,K.,  
 Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P., Venter,J.C.  
 Use of human BAC End Sequences for Sequence-Ready Map Building

**TITLE** Unpublished (1998)  
**JOURNAL** Other-GSS: RPC11-57FL.TJB  
**COMMENT** Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are derived from the human BAC library RPC1-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@jeong.med.buffalo.edu). Clones may be purchased from  
 BACPac Resources (http://bacpac.med.buffalo.edu/ordering) or from  
 Research Genetics (info@resgen.com). BAC end search page:  
 http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html  
 Class: BAC ends.

**FEATURES**  
**source**  
 1..119  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7521624"  
 /db\_xref="taxon:9606"  
 /clone="RPC1-11-57F1"  
 /clone\_lib="RPC1-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBACg3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPC111 Human Male BAC library"  
 RPc111 Human Male BAC library"

**BASE COUNT**  
 40 a 28 c 19 g 32 t

**ORIGIN**

Query Match 1.2%; Score 20; DB 17; Length 119;  
 Best local Similarity 100.0%; Pred. No. 93;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 1291 CTCCTTTGACTCAACTAA 1310  
 |||||||||||||||||||

**Db** 16 CTCCTTTGACTCAACTAA 35

**RESULT** 14  
**LOCUS** B0610731 254 bp mRNA linear EST 26-JUN-2002  
**DEFINITION** saps0605.y1 Gm-c1087 glycine max cDNA clone SOYBEAN CLONE ID:  
 Gm-c1087-3873 5', mRNA sequence.  
**ACCESSION** B0610731  
**VERSION** B0610731.1 GI:21600400  
**KEYWORDS** EST.  
**SOURCE** soybean.  
**ORGANISM** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.

**REFERENCE** 1 (bases 1 to 254)  
**AUTHORS** Shoemaker,R., Keim,P., Vodkin,L., Erpeiding,J., Coryell,V., Khanna  
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
 Wylie,T., Underwood,K., Steptoe,M., Theisler,B., Allen,M., Bowers  
 R., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
 J., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
 R., Waterston,R. and Wilson,R.

**TITLE** Public Soybean EST Project  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Resgen, Invitrogen Corp. 2130  
 South Memorial Parkway Huntville, AL 35801 For further information  
 call: (800)-533-4363 or contact: cc@resgen.com web site:  
 www.resgen.com

**FEATURES**  
**source**  
 1..254  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="SOYBEAN CLONE ID: Gm-c1087-3873"  
 /clone\_lib="Gm-c1087"  
 /tissue\_type="Roots"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:  
 XhoI; The mRNA was prepared using polyattract mRNA system  
 from PROMEGA. The cDNA was prepared using the STRATAGENE  
 kit. Complementary DNA was synthesized from mRNA using a  
 primer consisting of a poly(dT) sequence with a XhoI  
 restriction site  
 (5'GACGACAGAGAGAGAGACACTGCTCGAGTTTCTTTTCTTTTCTTTT).  
 EcoRI adapters (5'OH-AATTGGCGACGAG and 3'GCCGTCCTCP) were  
 ligated to the blunt-ended cDNA fragments followed by XhoI  
 digestion. The cDNA fragments were directionally cloned  
 into the EcoRI-XhoI restriction sites of the pluescript  
 vector. The vector was previously dephosphorylated. The  
 ligated cDNA fragments were transformed into DH10B by  
 electroporation. Library is in LB medium with ampicillin  
 and glycerol 8%. Average insert size: 800 bp."

**BASE COUNT**  
 81 a 42 c 47 g 84 t

**ORIGIN**

Query Match 1.2%; Score 20; DB 14; Length 254;  
 Best local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 115 ACGAGCTAAGAAAGTTCTG 134  
 |||||||||||||||||||

**Db** 186 ACGAGCTAAGAAAGTTCTG 167

**RESULT** 15  
**LOCUS** B0436799 289 bp mRNA linear EST 02-AUG-2000  
**DEFINITION** B0436799 RIKEN full-length enriched, adult pancreas islet cells Mus  
 musculus cDNA clone C82020F15 3', mRNA sequence.  
**ACCESSION** B0436799  
**VERSION** B0436799.1 GI:9276526  
**KEYWORDS** EST.  
**SOURCE** house mouse.  
**ORGANISM** Mus musculus

**REFERENCE** 1 (bases 1 to 289)  
**AUTHORS** Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
 P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
 Hirozane,T., Hori,F., Ishii,T., Ishikawa,U., Ishikawa,T., Itoh,M.,  
 Izawa,M., Kadota,K., Kagawa,T., Kai,C., Kawai,J., Kikuchi,N.,  
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
 Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
 Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,  
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya  
 T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamataka,I.,  
 Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
 M., Muramatsu,M. and Hayashizaki,Y.

**TITLE** RIKEN Mouse ESTs (Kono,H., et al.)



TITLE  
JOURNAL  
COMMENT

,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermostabilization and thermoinactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source

Location/Qualifiers  
1..303  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A530037K10"  
/clone\_lib="RIKEN full-length enriched, adult male aorta  
and vein"  
/sex="male"  
/tissue\_type="aorta and vein"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to rot = 20.0 and subtraction to rot = 459.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATCTCGAGTTAATTAAATTAATCCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I."

BASE COUNT 78 a 70 c 46 g 109 t  
ORIGIN  
Query Match 1.2%; Score 20; DB 10; Length 303;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1336 AAGAGCAAAAGTCAAGT 1355  
Db 120 AAGAGCAAAAGTCAAGT 101  
RESULT 18  
LOCUS BB381288 312 bp mRNA linear EST 13-JUL-2000  
DEFINITION BB381288 RIKEN full-length enriched, 0 day neonate cerebellum Mus  
musculus cDNA clone C230012B13 3', mRNA sequence.  
ACCESSION BB381288

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

BB381288.1 GI:9101012  
EST  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 312)  
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci  
,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,  
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,  
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,  
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,  
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,  
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata  
,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,  
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomihaga,N., Toya  
,T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,  
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino  
,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermostabilization and thermoinactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source

Location/Qualifiers  
1..312  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="C230012B13"  
/clone\_lib="RIKEN full-length enriched, 0 day neonate  
cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="0 day neonate"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to rot = 20.0 and subtraction to rot = 479.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATCTCGAGTTAATTAAATTAATCCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I."

BASE COUNT 103 a 45 c 53 g 111 t

ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 312;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 566 TAATGAATTTACAAATTC 585  
|||||  
Db 171 TAATGAATTTACAAATTC 190

RESULT 19  
AA712844 328 bp mRNA linear EST 24-DEC-1997  
LOCUS 32403 Lambda-PRU2 Arabidopsis thaliana cDNA clone 201N10T7, mRNA  
DEFINITION  
sequence.  
ACCESSION AA712844  
KEYWORDS AA712844.1 GI:2722761  
SOURCE EST.  
ORGANISM thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 328)  
AUTHORS Newman,T., debruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh  
,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel  
,E. and Somerville,C.  
Genes galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cDNA clones  
Plant. Physiol. 106, 1241-1255 (1994)  
JOURNAL MEDLINE  
COMMENT 95148729  
Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313cne@lm.cl.msu.edu  
Seq primer: T7 dye primer.  
FEATURES  
Source Location/Qualifiers  
1..328  
/organism="Arabidopsis thaliana"  
/strain="var Columbia"  
/db\_xref="taxon:3702"  
/clone="201N10T7"  
/clone\_lib="Lambda-PRU2"  
/note="Vector: Lambda Zip-Lox; Site\_1: Sal; Site\_2: Not;  
Lambda PRU2 is a cDNA library derived from equal  
quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
same plants as 3 but aerial tissue (stems, flowers and  
siliques. The vector is BRL's Lambda Zip-Lox. The cDNA  
inserts were directionally cloned with Sal-Not arms using  
oligo dt primed cDNA. "

BASE COUNT 80 a 63 c 81 g 91 t 13 others

ORIGIN

Query Match 1.2%; Score 20; DB 9; Length 338;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1308 AAAGAAATGCCTATATCCC 1327  
|||||  
Db 80 AAAGAAATGCCTATATCCC 99

RESULT 20  
AO697744 334 bp DNA linear GSS 06-JUL-1999  
LOCUS

DEFINITION HS-5553\_A2-D03.SP6E RPCI-11 Human Male BAC library Homo sapiens  
genomic clone plate-1129 Col-6 Row-G, DNA sequence.  
ACCESSION AO697744  
VERSION AO697744.1 GI:5387992  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 334)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
JOURNAL MEDLINE  
COMMENT 99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Resear h Genetics (Info@resgen.com). BAC end Web Server:  
http://www.hnsc.washington.edu  
Plate: 1129 row: G column: 6  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 334.  
FEATURES  
Source Location/Qualifiers  
1..334  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate-1129 Col-6 Row-G"  
/clone\_lib="RPCI-11 Human Male BAC library"  
/sex="male"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"

BASE COUNT 127 a 53 c 63 g 91 t

ORIGIN

Query Match 1.2%; Score 20; DB 17; Length 334;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AAATPAAAACTATCAGAT 21  
|||||  
Db 115 AAATPAAAACTATCAGAT 134

RESULT 21  
AZ016529/c 337 bp DNA linear GSS 25-FEB-2000  
LOCUS  
DEFINITION RPCI-23-368M9.TJ RPCI-23 Mus musculus genomic clone RPCI-23-368M9,  
DNA sequence.  
ACCESSION AZ016529  
VERSION AZ016529.1 GI:7091913  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 337)  
AUTHORS Zhao,S., Nierman,W., Felblyum,T., Malek,J., Shatsman,S., Akliuret  
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.

TITLE and Fraser,C.M.  
JOURNAL Mouse BAC End Sequences from Library RPCI-23  
COMMENT Unpublished (1999)  
Other\_CSSS: RPCI-23-368M9.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@ligr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@edlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.ligr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 368 row: M column: 9  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

## source

Location/Qualifiers  
1..337

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="RPCI-23-368M9"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 135 a 59 c 41 g 102 t  
ORIGIN

## Query Match

Best Local Similarity 1.2%; Score 20; DB 17; Length 337;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 ATTATCTTATGCTTATTTA 295  
|||||

DB 323 ATTATCTTATGCTTATTTA 304

RESULT 22 BEA32018 339 bp mRNA linear EST 18-MAY-2001  
LOCUS BEA32018.2 339 bp mRNA linear EST 18-MAY-2001  
DEFINITION EST398547 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA  
clone CLE64L20, mRNA sequence.

ACCESSION BEA32018  
VERSION BEA32018.2 GI:11387655  
KEYWORDS EST  
SOURCE tomato.

## ORGANISM

Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; eunasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.

REFERENCE 1 (bases 1 to 339)  
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,  
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,  
Niernan,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley  
,S.D.

TITLE Generation of ESTs from tomato fruit tissue, breaker stage  
JOURNAL Unpublished (2000)  
COMMENT On Jul 24, 2000 this sequence version replaced gi:9429861.  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University

100 Jordan Hall, Clemson, SC 29634, USA  
Email: http://www.genome.clemson.edu/orders/index.html  
5 prime sequence.

## FEATURES

## source

Location/Qualifiers  
1..339

/organism="Lycopersicon esculentum"  
/cultivar="FA496"  
/db\_xref="taxon:4081"  
/clone="CLE64L20"  
/clone\_lib="tomato breaker fruit, TIGR"  
/tissue\_type="pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/note="Vector: pBluescriptSMCmutadapt; Site\_1: EcoRI;  
Site\_2: XhoI; Fruit were harvested at the breaker stage  
(first sign of lycopene accumulation on the blossom end of  
the fruit). Fruit were cut in half and the seeds and  
locules were discarded prior to freezing the pericarp."

BASE COUNT 101 a 71 c 50 g 117 t  
ORIGIN

## Query Match

Best Local Similarity 1.2%; Score 20; DB 10; Length 339;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1571 AGAAGATTGACGTCCTGCTT 1590  
|||||

DB 65 AGAAGATTGACGTCCTGCTT 46

RESULT 23 A0685833 342 bp DNA linear GSS 28-JUN-1999  
LOCUS HS\_5549.AL.G08-SP6E RPCI-11 Human Male BAC Library Homo sapiens  
DEFINITION genomic clone Plate-1125 Col-15 Row-M, DNA sequence.  
ACCESSION A0685833  
VERSION A0685833.1 GI:5261904  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE 1 (bases 1 to 342)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@edlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Resea ch Genetics (info@resgen.com). BAC end web Server: http://www.htsc.washington.edu  
Plate: 1125 row: M column: 15  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 342.

## FEATURES

## source

Location/Qualifiers  
1..342

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-1125 Col-15 Row-M"  
/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACE3.6 vector at EcoRI sites"

BASE COUNT 118 a 55 c 58 g 111 t  
ORIGIN

Query Match 1.2%; Score 20; DB 17; Length 342;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 AATGTTCCATATGTTCTTCT 156  
|||||  
DB 137 AATGTTCCATATGTTCTTCT 156

RESULT 24 352 bp mRNA linear EST 19-APR-2001  
BG628369  
LOCUS cc-estficlel21p10a1 Tomato flower library from a mixture of  
developmental stages Lycopersicon esculentum cDNA clone  
cc-estficlel21p10a1, mRNA sequence.

ACCESSION BG628369 GI:13679842  
VERSION  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 352)  
van der Hoeven, R.S. and Tanksley, S.D.  
ESTs from a tomato flower library  
Unpublished (2001)  
Contact: Rutgers S. van der Hoeven  
Cornell University  
252 Emerson Hall, Ithaca, NY 14850, USA  
Tel: 607 255 7886  
Fax: 607 255 6683  
Email: rv19@cornell.edu  
3 prime sequence.

FEATURES  
source  
Location/Qualifiers  
1..352  
/organism="Lycopersicon esculentum"  
/cultivar="E6203"  
/db\_xref="taxon:4081"  
/clone="cc-estficlel21p10a1"  
/clone\_1lb="Tomato flower library from a mixture of  
developmental stages"  
/tissue\_type="developing flower buds and open flowers"  
/dev\_stage="4-8 week old plants"  
/lab\_host="X10LR"  
/note="Vector: pBK\_CMV; Site\_1: EcoRI; Site\_2: XhoI;  
Flowers and flower buds were collected from greenhouse  
grown plants and used for library construction (cTel)."

BASE COUNT 102 a 55 c 68 g 127 t  
ORIGIN

Query Match 1.2%; Score 20; DB 12; Length 352;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1571 AGAAGATTCAGCTCCTGCTT 1590  
|||||  
DB 252 AGAAGATTCAGCTCCTGCTT 271

RESULT 25 356 bp mRNA linear EST 19-APR-2001  
BG626629  
LOCUS

DEFINITION cc-estficlel3109a1 Tomato flower library from a mixture of  
developmental stages Lycopersicon esculentum cDNA clone  
cc-estficlel3109a1, mRNA sequence.

ACCESSION BG626629 GI:13678102  
VERSION  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 356)  
van der Hoeven, R.S. and Tanksley, S.D.  
ESTs from a tomato flower library  
Unpublished (2001)  
Contact: Rutgers S. van der Hoeven  
Cornell University  
252 Emerson Hall, Ithaca, NY 14850, USA  
Tel: 607 255 7886  
Fax: 607 255 6683  
Email: rv19@cornell.edu  
3 prime sequence.

FEATURES  
source  
Location/Qualifiers  
1..356  
/organism="Lycopersicon esculentum"  
/cultivar="E6203"  
/db\_xref="taxon:4081"  
/clone="cc-estficlel3109a1"  
/clone\_1lb="Tomato flower library from a mixture of  
developmental stages"  
/tissue\_type="developing flower buds and open flowers"  
/dev\_stage="4-8 week old plants"  
/lab\_host="X10LR"  
/note="Vector: pBK\_CMV; Site\_1: EcoRI; Site\_2: XhoI;  
Flowers and flower buds were collected from greenhouse  
grown plants and used for library construction (cTel)."

BASE COUNT 103 a 54 c 74 g 125 t  
ORIGIN

Query Match 1.2%; Score 20; DB 12; Length 356;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1571 AGAAGATTCAGCTCCTGCTT 1590  
|||||  
DB 261 AGAAGATTCAGCTCCTGCTT 280

RESULT 26 368 bp mRNA linear EST 19-APR-2001  
BG628526  
LOCUS cc-estficlel22116d1 Tomato flower library from a mixture of  
developmental stages Lycopersicon esculentum cDNA clone  
cc-estficlel22116d1, mRNA sequence.

ACCESSION BG628526 GI:13679999  
VERSION  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 368)  
van der Hoeven, R.S. and Tanksley, S.D.  
ESTs from a tomato flower library  
Unpublished (2001)  
Contact: Rutgers S. van der Hoeven  
Cornell University  
252 Emerson Hall, Ithaca, NY 14850, USA  
Tel: 607 255 7886  
Fax: 607 255 6683

Email: rv19@cornell.edu  
3 prime sequence.

FEATURES  
source

Location/Qualifiers  
1. 368  
/organism="Lycopersicon esculentum"  
/cultivar="E6203"  
/db\_xref="taxon:4081"  
/clone="CC-essfl:CLEL22116d1"  
/clone\_1lb="Tomato flower library from a mixture of developmental stages"  
/tissue\_type="developing flower buds and open flowers"  
/dev\_stage="4-8 week old plants"  
/lab\_host="XL0LR"  
/note="Vector: PBK CMV; Site 1: EcoRI; Site 2: XhoI; flowers and flower buds were collected from greenhouse grown plants and used for library construction (CLEL)."

BASE COUNT 106 a 56 c 75 g 131 t  
ORIGIN

Query Match 1.2%; Score 20; DB 12; Length 368;  
Best Local Similarity 100.0%; Pred. No. 1.le+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1571 AGAAGATTCAGCTCTCGCTT 1590  
|||||  
Db 262 AGAAGATTCAGCTCTCGCTT 281

RESULT 27  
AM011059 388 bp mRNA linear EST 10-SEP-1999  
LOCUS  
DEFINITION  
S716C01 Pine Triplex shoot tip library Pinus taeda cDNA clone  
S716C01, mRNA sequence.

ACCESSION AM011059  
VERSION AM011059.1 GI:5859837  
KEYWORDS EST.  
SOURCE loblolly pine.  
ORGANISM Pinus taeda

REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
AUTHORS Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.  
TITLE The Pine Gene Discovery Project  
JOURNAL Unpublished (1999)  
COMMENT Contact: Ross Whetten  
Forest Biotechnology Group  
North Carolina State University  
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh  
NC, 27695-8008  
Tel: 919-515-7800  
Fax: 919-515-7801  
Email: rosswhetten@ncsu.edu

Seq primer: 5' Lambda Triplex2 Sequencing Primer.  
Location/Qualifiers

FEATURES  
source

1. 388  
/organism="Pinus taeda"  
/db\_xref="taxon:3352"  
/clone="S716C01"  
/clone\_1lb="Pine Triplex shoot tip library"  
/lab\_host="E. coli BM25.8"  
/note="Organ: shoot tips; Vector: Lambda Triplex; Site 1: SfiI (A); Site 2: SfiI (B); Shoot tips (approx. 2 cm from apex) were collected during the spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the Lambda Triplex vector. Plasmid subclones in pTriplex were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

BASE COUNT 101 a 62 c 92 g 126 t 7 others  
ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 388;  
Best Local Similarity 100.0%; Pred. No. 1.le+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1494 ATTTCGCTGCTTCTGCAC 1513  
|||||  
Db 123 ATTGGTTGTTCTGCAC 142

RESULT 28  
A2630188 439 bp DNA linear GSS 13-DEC-2000  
LOCUS  
DEFINITION  
M0483D12R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0483D12 R, DNA sequence.  
A2630188  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 439)  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0483 row: D column: 12  
Seq primer: CACACAGAAACACCTATGACC  
Class: plasmid ends  
High quality sequence stop: 439.  
Location/Qualifiers

FEATURES  
source

1. 439  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0483D12"  
/clone\_1lb="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD2 (g1147321141gblAF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 140 a 69 c 53 g 177 t  
ORIGIN

Query Match 1.2%; Score 20; DB 17; Length 439;  
Best Local Similarity 100.0%; Pred. No. 1.le+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 851 TGCCTAACCTAAATGATA 870  
|||||  
Db 13 TGCCTAACCTAAATGATA 32

RESULT 29  
AM218715/c 440 bp mRNA linear EST 18-MAY-2001  
LOCUS EST301195 tomato root during/after fruit set, Cornell University  
DEFINITION Lycopersicon esculentum cDNA clone cLEX1A3, mRNA sequence.  
ACCESSION AM218715  
VERSION AM218715.1 GI:6529589  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 440)  
van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Liang,F.,  
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Romning,C.M.,  
Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.  
Generation of ESTs from tomato root tissue  
Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
Location/Qualifiers  
1..440  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEX1A3"  
/clone\_1lb="tomato root during/after fruit set, Cornell  
University"  
/tissue\_type="root"  
/dev\_stage="plants during and after fruit-set"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Tanksley; Tissue supplied by Dave Garvin  
(USDA-ARS, Ithaca, NY 14850)."  
BASE COUNT 133 a 92 c 71 g 144 t  
ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 440;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1571 AGAAGATTACGCTCCTT 1590  
|||||  
Db 166 AGAAGATTACGCTCCTT 147

RESULT 30  
AM218716 440 bp mRNA linear EST 18-MAY-2001  
LOCUS EST301196 tomato root during/after fruit set, Cornell University  
DEFINITION Lycopersicon esculentum cDNA clone cLEX1A3, mRNA sequence.  
ACCESSION AM218716  
VERSION AM218716.1 GI:6529590  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 440)

AUTHORS van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Liang,F.,  
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Romning,C.M.,  
Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.  
Generation of ESTs from tomato root tissue  
Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
3 prime sequence.  
Location/Qualifiers  
1..440  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEX1A3"  
/clone\_1lb="tomato root during/after fruit set, Cornell  
University"  
/tissue\_type="root"  
/dev\_stage="plants during and after fruit-set"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Tanksley; Tissue supplied by Dave Garvin  
(USDA-ARS, Ithaca, NY 14850)."  
BASE COUNT 144 a 71 c 92 g 133 t  
ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 440;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1571 AGAAGATTACGCTCCTT 1590  
|||||  
Db 275 AGAAGATTACGCTCCTT 294

RESULT 31  
AM933403/c 458 bp mRNA linear EST 18-MAY-2001  
LOCUS EST359342 tomato fruit mature green, TAMU Lycopersicon esculentum  
DEFINITION cDNA clone cLEF54010 5', mRNA sequence.  
ACCESSION AM933403  
VERSION AM933403.1 GI:8108900  
KEYWORDS EST.  
SOURCE  
ORGANISM  
Lycopersicon esculentum  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 458)  
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,  
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Romning,C.M.,  
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.J.  
Generation of ESTs from tomato fruit tissue  
Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
Location/Qualifiers  
1..458  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="cLEF54010"  
/clone\_1lb="tomato fruit mature green, TAMU"  
/tissue\_type="fruit pericarp"  
/dev\_stage="mature green (3-5 days pre-ripening)"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:



XhoI: CLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'Immature' and the seeds and locules were discarded prior to freezing the pericarp"

BASE COUNT 141 a 103 c 99 g 115 t  
ORIGIN

Query Match 1.2%; Score 20; DB 10; Length 458;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAGATTCAGCTCCTGCTT 1590  
|||||  
Db 381 AGAGATTCAGCTCCTGCTT 362

RESULT 32  
AQ433974 472 bp DNA linear GSS 31-MAR-1999  
LOCUS

DEFINITION HS.5088.AL.D04.SP6E RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate-664 Col-7 Row-G, DNA sequence.

ACCESSION AQ433974  
VERSION AQ433974.1 GI:4544309  
KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 472)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)  
or from Research Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu  
Plate: 664 Row: G Column: 7  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 472.

FEATURES  
source  
1..472  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=664 Col=7 Row=G"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACe3.6 vector at EcoRI sites"

BASE COUNT 166 a 75 c 91 g 139 t 1 others  
ORIGIN

Query Match 1.2%; Score 20; DB 17; Length 472;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AAATAAAACTATCAGAT 21  
|||||  
Db 120 AAATAAAACTATCAGAT 139

RESULT 33  
BF098187/c 496 bp mRNA linear EST 18-MAY-2001  
LOCUS

DEFINITION EST428796 tomato nutrient deficient roots Lycopersicon esculentum  
cDNA clone cLEM26C18 5' sequence, mRNA sequence.

ACCESSION BF098187  
VERSION BF098187.1 GI:10903985  
KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 496)  
van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,F.,  
Upton,J., Hansen,T.S., Rohning,C.M., Craven,M.B., Bowman,C.L.,  
Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.  
and Tanksley,S.D.  
Generation of ESTs from tomato nutrient deficient roots  
unpublished (1999)

TITLE Contact: CUGI  
JOURNAL Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: http://www.genome.clemson.edu/orders/index.html.

COMMENT Location/Qualifiers  
1..496  
/organism="Lycopersicon esculentum"  
/cultivar="TA492"  
/db\_xref="taxon:4081"  
/clone="cLEM26C18"  
/clone\_lib="tomato nutrient deficient roots"  
/tissue\_type="Roots"  
/dev\_stage="5-6 weeks old"  
/lab\_host="SOLR"

/note="Vector: pBluescriptSMCudaapt; Site\_1: 5' EcoRI;  
Site\_2: 3' XhoI; Roots were harvested from plants grown  
under the following deficiencies/stresses: 10 mM Al, Zn,  
P, K, Fe, N. mRNA was isolated from individual treatments.  
Proportional aliquots of mRNA of each treatment were mixed  
and used for library construction."

BASE COUNT 149 a 115 c 117 g 115 t  
ORIGIN

Query Match 1.2%; Score 20; DB 12; Length 496;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1571 AGAGATTCAGCTCCTGCTT 1590  
|||||  
Db 469 AGAGATTCAGCTCCTGCTT 450

RESULT 34  
AM651372/c 498 bp mRNA linear EST 18-MAY-2001  
LOCUS

DEFINITION EST329826 tomato germinating seedlings, TAMO Lycopersicon  
esculentum cDNA clone CLE116M22 5', mRNA sequence.

ACCESSION AM651372  
VERSION AM651372.1 GI:7412610  
KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 498)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterales; eusterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.

```

REFERENCE
AUTHORS      1 (bases 1 to 498)
              Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S. Holt,I.E.,
              Liang,F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M.,
              Nierman,W., Fraser,C.M., Giovannoni,J.J., Martin,G.B. and Tanksley
              ,S.D.
TITLE        Generation of ESTs from germinating tomato seed
JOURNAL      Unpublished (2000)
COMMENT      Contact: CUGI
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Email: http://www.genome.clemson.edu/orders/index.html
              5 prime sequence.
FEATURES
SOURCE
              Location/Qualifiers
              1..498
                /organism="Lycopersicon esculentum"
                /cultivar="TA496"
                /db_xref="taxon:4081"
                /clone_id="CLEF16M2"
                /clone_lib="tomato germinating seedlings, TAMU"
                /tissue_type="whole seedlings"
                /dev_stage="7 days post imbibition"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
              XhoI; 7 days post imbibition on water-agar. Mixed stage
              whole germinating seedlings from seed coat emergence up
              to two centimeters in seeds not showing obvious signs of
              germination were discarded."
BASE COUNT   149 a      112 c      123 g      114 t
ORIGIN
Query Match      1.2%; Score 20; DB 10; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1571 AGAAGATTACGCTCTGCTT 1590
      |||||||
Db  493 AGAAGATTACGCTCTGCTT 474

RESULT 35
BM404442/c 500 bp mRNA linear EST 22-JAN-2002
LOCUS      EST578769 potato roots Solanum tuberosum cDNA cPRO20D24 5'
DEFINITION end, mRNA sequence.
ACCESSION  BM404442
VERSION     BM404442.1 GI:18255884
KEYWORDS   EST.
SOURCE      potato.
ORGANISM    Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
            1 (bases 1 to 500)
REFERENCE   van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S.,
            Utechtack,T., Chiening,A., Bougri,O., Buell,C.R., Konning,C.,
            Tanksley,S. and Baker,B.
            Generation of ESTs from potato roots
            Unpublished (2001)
TITLE      Contact: Research Genetics, Libraries Division
JOURNAL    Tel: 1-800-711-6195
            Email: cdna@resgen.com
COMMENT    For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: T3.
FEATURES
SOURCE      Location/Qualifiers
              1..500
                /organism="Solanum tuberosum"
                /cultivar="Kennebec"
                /db_xref="taxon:4113"
                /clone="cPRO20D24"
                /clone_lib="potato roots"
                /tissue_type="roots"
                /dev_stage="in vitro grown stem cuttings"

```

```

/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Supplier: Cornell University, Tanksley lab;
Sequencing: The Institute for Genomic Research. Roots were
isolated from in vitro grown stem cuttings on CM medium.
Roots were isolated two weeks after placing the stem
cuttings from in vitro grown plants on medium."
BASE COUNT   142 a      116 c      129 g      113 t
ORIGIN
Query Match      1.2%; Score 20; DB 13; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1571 AGAAGATTACGCTCTGCTT 1590
      |||||||
Db  493 AGAAGATTACGCTCTGCTT 474

RESULT 36
AM221154/c 514 bp mRNA linear EST 18-MAY-2001
LOCUS      EST297623 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone CLEF13A10, mRNA sequence.
ACCESSION  AM221154
VERSION     AM221154.1 GI:6532838
KEYWORDS   EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
            1 (bases 1 to 514)
REFERENCE   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ann,S., Ronning,C.M.,
            Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
            Generation of ESTs from tomato fruit tissue
            Unpublished (1999)
TITLE      Contact: CUGI
JOURNAL    Clemson University Genomics Institute
COMMENT    Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES
SOURCE      Location/Qualifiers
              1..514
                /organism="Lycopersicon esculentum"
                /cultivar="TA496"
                /db_xref="taxon:4081"
                /clone="CLEF13A10"
                /clone_lib="tomato fruit mature green, TAMU"
                /tissue_type="fruit pericarp"
                /dev_stage="mature green (3-5 days pre-ripening)"
                /lab_host="SOLR"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
              XhoI; CLEF - Fruit were tagged at the 1cm stage and
              harvested 3-5 days prior to ripening. Fruit were cut in
              half to verify the seeds were indeed 'immature' and the
              seeds and locules were discarded prior to freezing the
              pericarp"
BASE COUNT   158 a      118 c      122 g      116 t
ORIGIN
Query Match      1.2%; Score 20; DB 10; Length 514;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1571 AGAAGATTACGCTCTGCTT 1590
      |||||||
Db  503 AGAAGATTACGCTCTGCTT 484

```

RESULT 37  
BB284364  
LOCUS  
DEFINITION BB284364 RIKEN full-length enriched, adult retina Mus musculus cDNA  
clone A930105K03 3', mRNA sequence.  
ACCESSION BB284364  
VERSION BB284364  
KEYWORDS  
SOURCE BB284364.1 GI:8984813  
EST.  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 546)  
AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Kono, H., et al.)  
TITLE Unpublished (2000)  
JOURNAL  
COMMENT Contact: Yoshitake Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome-gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.  
FEATURES  
Location/Qualifiers  
1..546  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A930105K03"  
/clone\_lib="RIKEN full-length enriched, adult retina"  
/tissue\_type="retina"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGAGATCCAAAGCTTTTCTTTTCTTTTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of

sequence [5'GAGAGAGAGATTCGAGTTAATTAATTAATTCACCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I. -Retina RNA was provided by Stefano Gustlinch,  
Department of Neurobiology, Harvard Medical School, 220  
Longwood Ave., Boston, MA02115, USA, whose assistance we  
gratefully acknowledge."  
BASE COUNT 180 a 112 c 103 g 151 t  
ORIGIN  
Query Match 1.2%; Score 20; DB 10; Length 546;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1333 ACCAAGAGCAAAAGTCANA 1352  
DB 328 ACCAAGAGCAAAAGTCANA 347  
RESULT 38  
LOCUS A0121410 548 bp DNA linear GSS 22-SEP-1998  
DEFINITION HS\_3087\_B2\_B12\_MF CTF Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=3087 Col=24 Row=D, DNA sequence.  
ACCESSION A0121410  
VERSION A0121410.1 GI:3498576  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 548)  
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
JOURNAL  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3087 row: D column: 24  
Class: BAC ends  
High quality sequence stop: 548.  
FEATURES  
Location/Qualifiers  
1..548  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=3087 Col=24 Row=D"  
/clone\_lib="CTF Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"  
BASE COUNT 191 a 93 c 109 g 149 t 6 others  
ORIGIN  
Query Match 1.2%; Score 20; DB 17; Length 548;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 AAATAAANAATCATCAGAT 21  
DB 167 AAATAAANAATCATCAGAT 186  
RESULT 39  
B0119131

LOCUS B0119131 558 bp mRNA linear EST 17-APR-2002  
 DEFINITION EST604707 mixed potato tissues Solanum tuberosum cDNA clone STMEH45  
 3' end, mRNA sequence.  
 ACCESSION B0119131  
 VERSION B0119131.1 GI:20171093  
 KEYWORDS EST.  
 SOURCE potato.  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; eusterids I; Solanales; Solanaceae; Solanum.  
 1 (bases 1 to 558)  
 Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,  
 Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and  
 Karameyeva,S.A.  
 Generation of a set of potato cDNA clones for microarray analyses  
 Unpublished (2002)  
 CONTACT Contact: Robin Buell  
 The Institute for Genomic Research  
 9712 Medical Center Dr, Rockville, MD 20850, USA  
 Email: potato@tigr.org  
 This clone is available through the Research Genetics, contact the  
 Research Genetics for further information 1-800-711-6195 or  
 cdna@resgen.com  
 Seq primer: 77.  
 FEATURES  
 source  
 1..558  
 /organism="Solanum tuberosum"  
 /cultivar="Kennebec or Binje"  
 /db\_xref="taxon:4113"  
 /clone="STMEH45"  
 /clone\_lib="mixed potato tissues"  
 /tissue\_type="mixed tissues"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:  
 XhoI; supplier: Combination of untreated and phytophthora  
 infestans-treated libraries of stolons, leaves, leaflets,  
 axillary buds of stem explants, petioles, germinating eyes  
 , tubers, or roots."  
 BASE COUNT 170 a 120 c 91 g 177 t  
 ORIGIN  
 Query Match 1.2%; Score 20; DB 14; Length 558;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1270 TTTGCAATCCGCAAAATAC 1289  
 ||||||||||||||||||  
 Db 18 TTTGCAATCCGCAAAATAC 37  
 RESULT 40  
 BE459902/c 559 bp mRNA linear EST 18-MAY-2001  
 LOCUS EST415194 tomato developing/immature green fruit Lycopersicon  
 DEFINITION esculentum cDNA clone CLEM8120, mRNA sequence.  
 ACCESSION BE459902  
 VERSION BE459902.1 GI:9504204  
 KEYWORDS EST.  
 SOURCE tomato.  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; eusterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.  
 1 (bases 1 to 559)  
 Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,  
 Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Romling,C.M.,  
 Niernan,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley  
 ,S.D.  
 Generation of ESTs from tomato fruit tissue, immature green  
 Unpublished (2000)  
 CONTACT Contact: CUGI  
 TITLE  
 JOURNAL  
 COMMENT

Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.  
 FEATURES  
 source  
 1..559  
 /organism="Lycopersicon esculentum"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="CLEM8120"  
 /clone\_lib="tomato developing/immature green fruit"  
 /tissue\_type="fruit"  
 /dev\_stage="immature green (5-35 days post-anthesis)"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescriptSMCudapt; Site.1: EcoRI;  
 Site.2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and  
 harvested at 7 day intervals through 35 dpa. Equal masses  
 of tissue from each stage were combined (including seeds  
 and locules) prior to mRNA isolation."  
 BASE COUNT 174 a 127 c 129 g 129 t  
 ORIGIN  
 Query Match 1.2%; Score 20; DB 10; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1571 AGAGATTTCAGCTCCTGCTT 1590  
 ||||||||||||||||||  
 Db 504 AGAGATTTCAGCTCCTGCTT 485  
 Search completed: January 21, 2003, 21:43:48  
 Job time : 1696 secs

Bb 404 / 4U AATTTCCTGACCGTTTAAAGGCCATCCCAACCCTCAGGGTAATCCCAGGC

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USPTO)**